VOLUME 106 PART B NUMBER 26 MARCH 1959



The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE · LONDON W.C.2

Price Ten Shillings and Sixpence

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1958–1959

President

S. E. GOODALL, M.Sc.(Eng.).

Past-Presidents

W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.C.M.G., K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D., LL.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
SIR JOSIAH ECCLES, C.B.E., D.Sc.
SIR GEORGE H. NELSON, Bart.
SIR GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).
T. E. GOLDUP, C.B.E.

Vice-Presidents

SIR WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S. G. S. C. LUCAS, O.B.E. SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
C. T. MELLING, C.B.E., M.Sc.Tech. A. H. MUMFORD, O.B.E., B.Sc.(Eng.).

Honorary Treasurer

E. LEETE.

Ordinary Members of Council

J. A. BROUGHALL, B.Sc.(Eng.).
PROFESSOR M. W. HUMPHREY DAVIES, M.Sc.
SIR JOHN DEAN, B.Sc.
B. DONKIN, B.A.
J. M. FERGUSON, B.Sc.(Eng.).
D. C. FLACK, B.Sc.(Eng.), Ph.D.
J. S. FORREST, D.Sc., M.A.
R. J. HALSEY, C.M.G., B.Sc.(Eng.).
E. M. HICKIN.
J. B. HIGHAM, Ph.D., B.Sc.
F. C. MCLEAN, C.B.E., B.Sc.
B. L. METCALF, B.Sc.(Eng.).
J. R. MORTLOCK, Ph.D., B.Sc.(Eng.).
R. H. PHILLIPS, T.D.
H. V. PUGH.
J. R. RYLANDS, M.Sc., J.P.
D. P. SAYERS, B.Sc.
C. E. STRONG, O.B.E., B.A., B.A.I.
D. H. TOMPSETT, B.Sc.(Eng.).
H. WATSON-JONES, M.Eng.
H. WEST, M.Sc.

Chairmen and Past-Chairmen of Sections

Measurement and Control:
J. K. WEBB, M.Sc.(Eng.), B.Sc.Tech.
*H. S. PETCH, B.Sc.(Eng.).

Electronics and Communications:
G. MILLINGTON, M.A., B.Sc.
*J. S. MCPETRIE, Ph.D., D.Sc.

Supply:
D. P. SAYERS, B.Sc.
*PROFESSOR M. G. SAY, Ph.D., M.Sc., F.R.S.E.

Utilization:
R. A. MARRYAT, B.Sc.(Eng.).
*J. VAUGHAN HARRIES.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
D. E. LAMBERT, B.Sc.(Eng.).
*J. D. PIERCE.

Mersey and North Wales Centre:
J. COLLINS.
*T. MAKIN.

North Midland Centre:
J. D. NICHOLSON, B.Sc.
*A. J. COVENEY.

North-Eastern Centre:
A. T. CRAWFORD, B.Sc.
*T. W. WILCOX.

Southern Centre:
G. BISHOP, B.Sc.
*L. G. A. SIMS, M.Sc., Ph.D., D.Sc.

North-Western Centre:
PROFESSOR F. C. WILLIAMS, O.B.E., D.Sc., D.Phil., F.R.S.
*F. R. PERRY, M.Sc.Tech.

Northern Ireland Centre:
D. S. MCILLHAGGER, Ph.D., M.Sc.
*C. M. STOUPE, B.Sc.

Western Centre:
R. W. STEEL.
*J. F. WRIGHT.

Scottish Centre:
R. J. RENNIE, B.Sc.
*E. O. TAYLOR, B.Sc., F.R.S.E.

South Midland Centre:
J. ASHMORE.
*L. L. TOLLEY, B.Sc.(Eng.).

* Past Chairman.

MEASUREMENT AND CONTROL SECTION COMMITTEE 1958–1959

Chairman

. K. WEBB, M.Sc.(Eng.), B.Sc.Tech.

Vice-Chairmen

PROFESSOR A. TUSTIN, M.Sc.; C. G. GARTON.

Past-Chairmen

H. S. PETCH, B.Sc.(Eng.); DENIS TAYLOR, M.Sc., Ph.D.

Ordinary Members of Committee

E. W. CONNON, B.Sc.(Eng.), M.Eng.
W. S. ELLIOTT, M.A.
W. C. LISTER, B.Sc.
A. C. LYNCH, M.A., B.Sc.
A. J. MADDOCK, D.Sc.
R. E. MARTIN.
R. S. MEDLOCK, B.Sc.
A. NEMET, Dr.Sc.Techn.
G. A. V. SOWTER, Ph.D., B.Sc.(Eng.).
R. H. TIZARD, B.A.
F. C. WIDDIS, B.Sc.(Eng.).
M. V. WILKES, M.A., Ph.D., F.R.S.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
J. B. HIGHAM, Ph.D., B.Sc. (representing the Council).
W. GRAY (representing the North-Eastern Electronics and Measurement Group).
T. KILBURN, M.A., Ph.D., D.Sc. (representing the North-Western Measurement and Control Group).
G. H. RAYNER, B.A. (nominated by the National Physical Laboratory).
H. M. GALE, B.Sc.(Eng.) (representing the South Midland Electronics and Measurement Group).

ELECTRONICS AND COMMUNICATIONS SECTION COMMITTEE 1958–1959

Chairman

G. MILLINGTON, M.A., B.Sc.

Vice-Chairmen

M. J. L. PULLING, C.B.E., M.A.
R. J. HALSEY, C.M.G., B.Sc.(Eng.).
T. B. D. TERRONI, B.Sc.

Past-Chairmen

J. S. MCPETRIE, Ph.D., D.Sc.
R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).

Ordinary Members of Committee

D. A. BARRON, M.Sc.
W. J. BRAY, M.Sc.(Eng.).
J. BROWN, M.A., Ph.D.
PROFESSOR A. L. CULLEN, Ph.D., B.Sc.
C. W. EARP, B.A.
L. I. FARREN, M.B.E.
V. J. FRANCIS, B.Sc.
G. G. GOURIET.
COMDR. C. G. MAYER, O.B.E., U.S.N.R.
J. MOIR.
B. G. PRESSEY, M.Sc.(Eng.), Ph.D.
T. R. SCOTT, D.F.C., B.Sc.
W. E. WILLSHAW, M.B.E., M.Sc.Tech.
A. J. YOUNG, B.Sc.(Eng.).

And

The President (*ex officio*).
The Chairman of the Papers Committee.
C. E. STRONG, O.B.E., B.A., B.A.I. (representing the Council).
E. H. COOKE-YARBOROUGH, M.A. (Co-opted Member).
G. H. HICKLING, B.Sc. (representing the North-Eastern Electronics and Measurement Group).
N. C. ROLFE, B.Sc.(Eng.) (representing the Cambridge Electronics and Communications Group).
K. J. BUTLER (representing the North-Western Electronics and Communications Group).
R. E. YOUNG, B.Sc. (Eng.) (representing the South Midland Electronics and Measurement Group).

The following nominees:
Royal Navy: CAPTAIN R. L. CLODE, R.N.
Army: COL. R. G. MILLER, M.A.
Royal Air Force: GROUP CAPTAIN A. FODEN, B.Sc.Tech., R.A.F.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

F. JERVIS SMITH, M.I.E.E.

Principal Assistant Secretary

F. C. HARRIS.

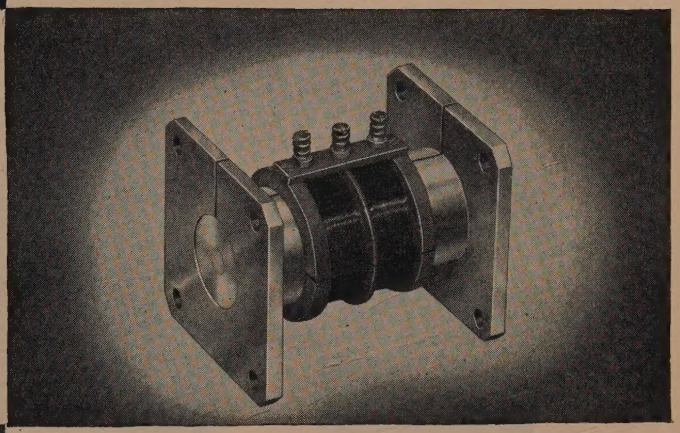
Editor-in-Chief

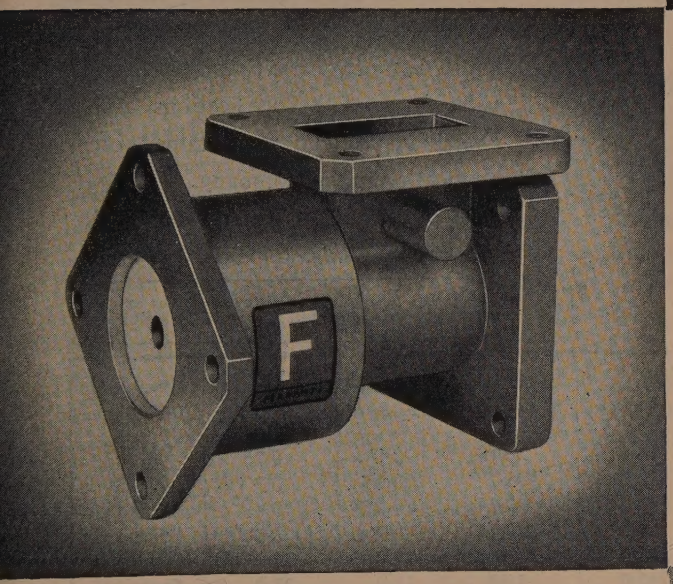
G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

FERRANTI X-BAND FERRITE DEVICES

FERRITE SWITCH TYPE 3F

Frequency 9600 to 9800 mc/s. Peak attenuation 30 db min. Insertion loss ·5 db max. Power handling capacity 30 watts. Weight 3 oz.



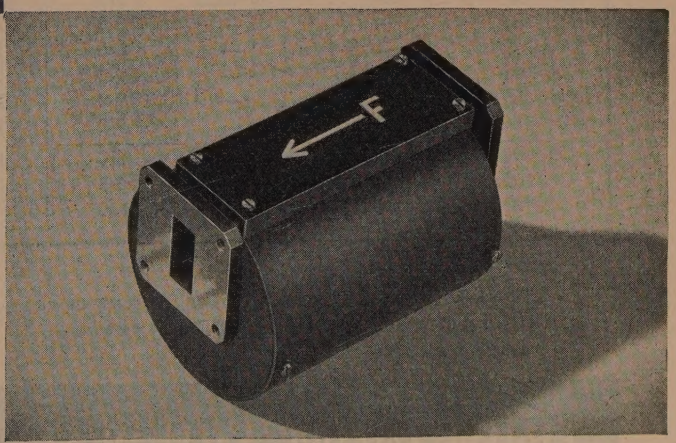


FERRITE CIRCULATOR

A compact three port circulator capable of handling 50 kW peak, 50 watt mean. It can be pre-set to cover a bandwidth of 400 mc/s in the 3 cm. band. The magnetic field is supplied by permanent magnets. The total weight of the component is approximately 4½ oz.

FERRITE ISOLATOR

Type 2F/2. 100 kW peak, 100 watt mean.
Type 2F/3. 200 kW peak, 200 watt mean.
Isolation in both cases greater than 20 db over a bandwidth of 1000 mc/s with an insertion loss of less than 1 db.

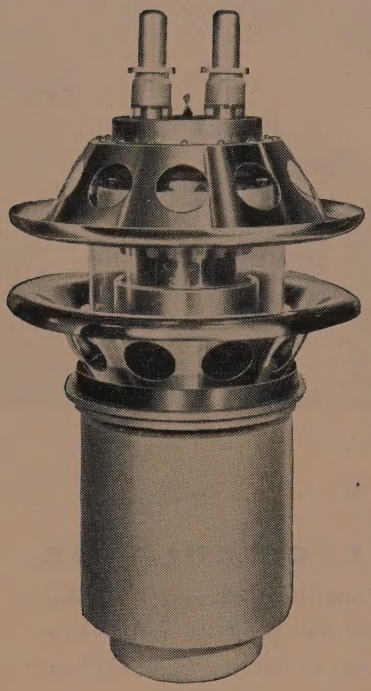


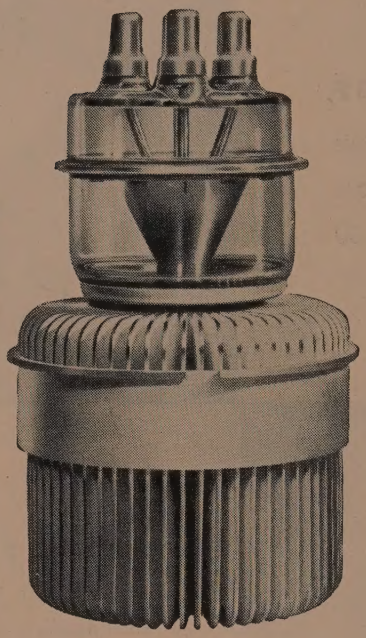
FERRANTI

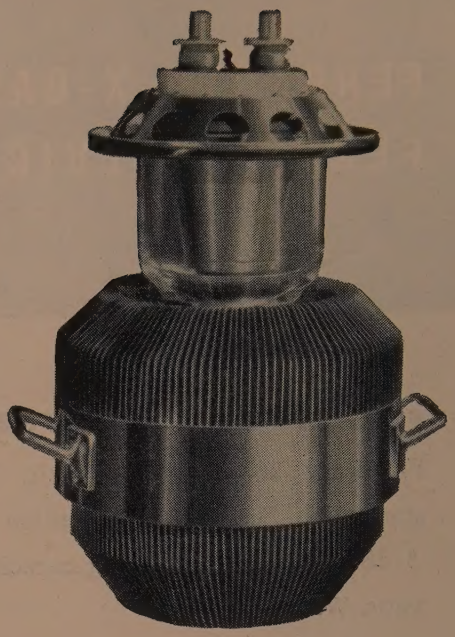
FERRANTI LTD · KINGS CROSS ROAD · DUNDEE Telephone: DUNDEE 87141

ES/T 52

TRANSMITTING VALVES







The tables below give data relating to some of the Transmitting Valves manufactured by the English Electric Valve Company. The full list of these valves, complete specifications and characteristics can be obtained on application to the Company.

All valves listed below have thoriated tungsten filaments.

Please visit our private display of valves at the Kensington Palace Hotel, March 17–21 inclusive.

E.E.V. Type	Equivalent Types	Service Type	Filament		Frequency MC/S §	Anode voltage max. (kV)	Anode dissipation max. (kW)	Type of cooling
			Voltage (V)	Current (A)				
5867	5867, TY3-250	CV1350	5·0	14·0	150	3·0	0·25	Natural to 30 Mc/s
833A	TY4-350, 833A	CV635	10·0	10·0	30	3·0 4·0	0·3 0·4	Natural Forced air
‡ 5762	TY6-5000A, ACT30	CV2383	12·6	29·0	30/220	6·2	3·0	Forced air
BR 161	—	CV2322	9·0	175·0	30/50	12·0	15·0	Forced air
BR 189	—	—	9·0	240·0	5/50	15·0	27·0	Forced air
BR 1122	—	—	6·0	115·0	5/110	12·0	10·0	Forced air
BW 161	—	—	9·0	175·0	30/50	12·0	30·0	Water
BW 189	—	—	9·0	240·0	5/50	15·0	35·0	Water
BY 189	—	—	9·0	240·0	5/50	15·0	35·0	Vapour
CR 192	6166	—	5·0	175·0	30/220	6·0	10·0	Forced air
CR1100	QY5-3000A	—	6·3	32·5	110/220	5·0	3·0	Forced air
6181	6181, CR1101	—	120	1·6	900	2·0	2·0	Forced air

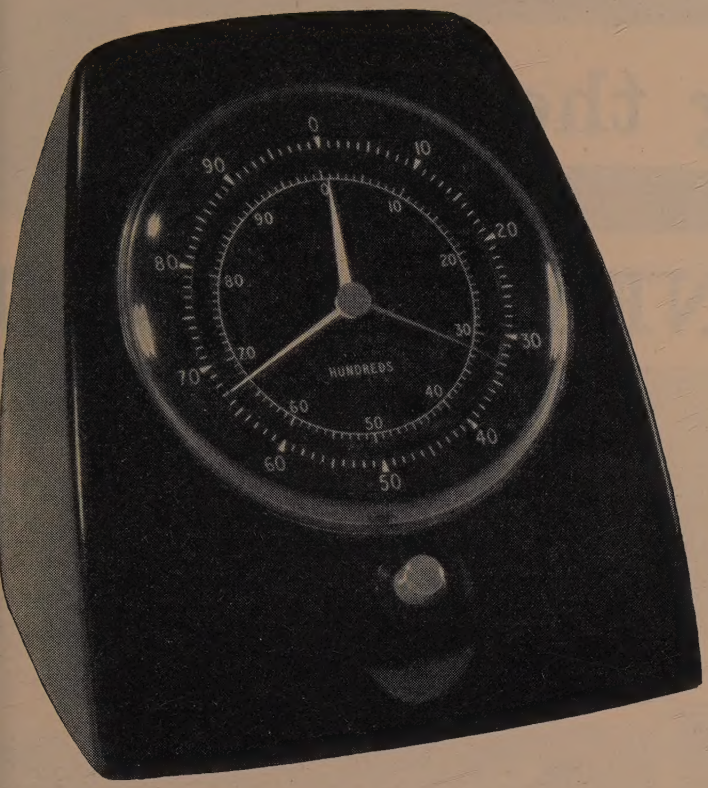
‡ *Previously BR 191 B*

§ *The lower value indicates the operating frequency at full rating. Operation at the higher value is possible with suitable derating.*

'ENGLISH ELECTRIC'

ENGLISH ELECTRIC VALVE CO. LTD. *Chelmsford, England*
Telephone: Chelmsford 3491

AP/123



BACKGROUND On 1st December, 1958, the G.P.O.'s new Subscriber Trunk Dialling (S.T.D.) scheme came into operation on the Bristol Central Exchange, by 1970 it will embrace three quarters of all trunk calls in the United Kingdom.

When S.T.D. is introduced, *all* telephone calls, whether trunk or local, can be dialled. The calls will be charged in twopenny units. *It is these units that SMITHS Subscriber's Private Meter has been designed to indicate.*

THE INSTRUMENT This consists essentially of a meter, designed to British G.P.O. specification, housed in a shock resisting plastic case. The meter dial has two scales marked 0-100 units. The outer scale in yellow registers units, the inner white scale registers hundreds. A yellow and white pointer are associated with the two scales.

A third pointer in red also registers units but can be reset to zero by depressing the red button in order to register the cost of an individual call. Operation of the reset button does not interrupt the function of the meter and the yellow and white pointers will continue to summate the call charges.

SALIENT FEATURES No audible noise interference.

Does not respond to voltages less than 100V. 25 c/s A.C. under normal ringing conditions or fault conditions.

This meter will be available to subscribers on the S.T.D. system in the U.K. by rental from the G.P.O.

Overseas enquiries should be made direct to the address below.

SMITHS

are proud to be associated with S.T.D.

SMITHS

INDUSTRIAL INSTRUMENT DIVISION

Chronos Works, North Circular Road, London, N.W.2. Telephone: GLAdstone 1136

Introducing the MC10

a NEW 5-Channel

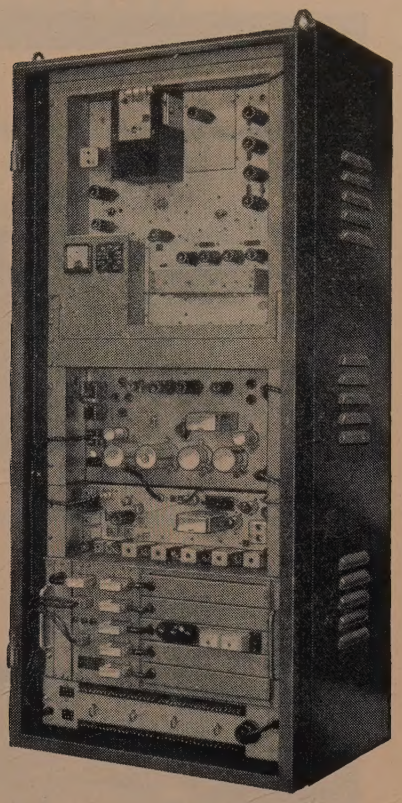
VHF Radio



A TYPICAL LOCATION

where this new 5-Channel Transistorised VHF Radio-Telephone Terminal will prove indispensable in vital communications

Transistorised Telephone Terminal



Compact design which at low cost provides five high-grade telephone circuits.

Radio

Frequency range 156 - 184 Mc/s
(other frequency ranges are available)
Transmitter power output 30 watts
Deviation 75 kc/s
Receiver Noise Factor 8db
All characteristics of the transmitter and receiver conform to CCIR specifications

Carrier Telephone Equipment

5 Telephone channels · 4 kc/s spaced
Equipped with Out of Band Signalling
Facilities for dialling, Ringdown or junction working
Printed Wiring · Plug-in Units
Crystal frequency control
Resin cast components

The result of co-operative enterprise between two great organizations

Redifon — **SIEMENS EDISWAN**

REDIFON LIMITED
Communications Division, Wandsworth, London, S.W.18
Telephone : VANdyke 7281
A Manufacturing Company in the Rediffusion Group

SIEMENS EDISON SWAN LIMITED
Transmission Division, Woolwich, London, S.E.18
Telephone : Woolwich 2020
An A.E.I. Company

Marconi in Broadcasting

80 countries of the world rely on Marconi broadcasting equipment

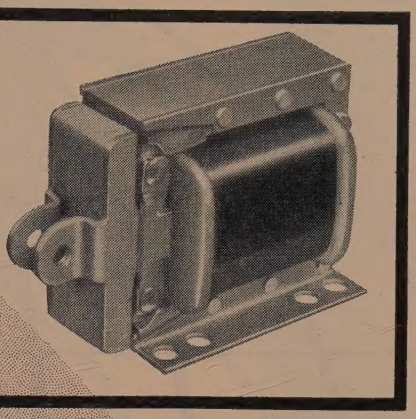


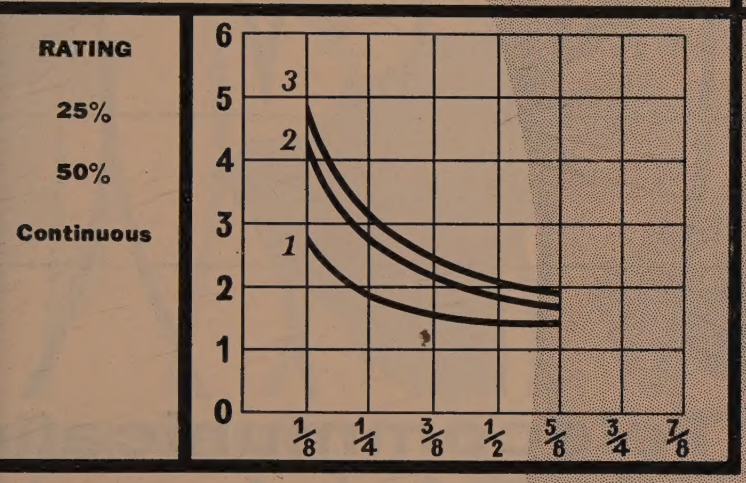
MARCONI COMPLETE SOUND BROADCASTING SYSTEMS

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX, ENGLAND

M3

A New Varley Solenoid . . .

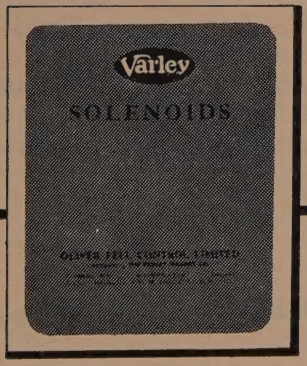




For Automatic Vending Machines

Varley introduce their new 'V' type solenoid specifically developed for application to automatic vending machines.

For latest details of the new Varley catalogue with the complete range of Varley solenoids, mail this coupon to:



OLIVER PELL CONTROL LIMITED
Cambridge Row, Woolwich, London, S.E.18.
Cables: Varleymag, Woolwich. Telephone: Woolwich 1422

Name..... Company.....

Address.....

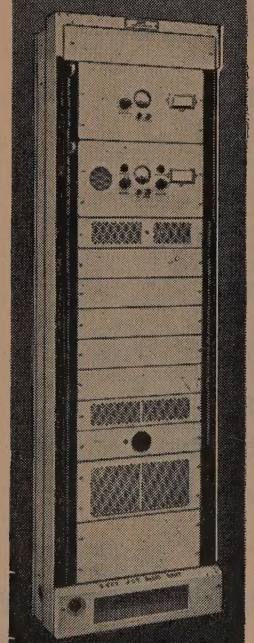
OPS 9

VHF



G.E.C.

Radio Telecommunicatio



5 circuit V.H.F
junction radio
terminal equipment.

G.E.C.

Over 1000 5-circuit V.H.F. Junction Radio Equipments have been supplied to 28 countries throughout the world.

The vast experience gained by G.E.C. in surveying, planning, designing, ma

EVERYTHING F

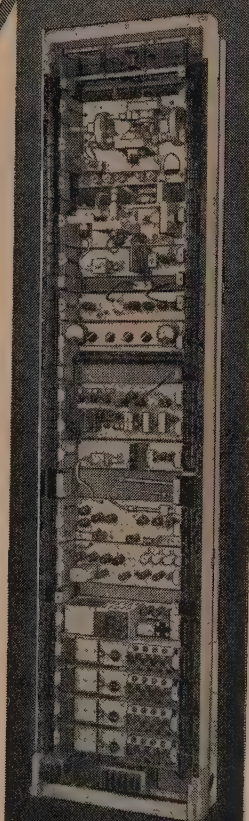
THE GENERAL ELECTRIC COMPANY LIMITED OF ENG



UHF

ıipment
throughout the world

Microwave systems providing over 4950 wide band channel miles for television or multi-circuit telephony have been installed or ordered for 7 countries throughout the world.



240-circuit U.H.F. radio terminal equipment (less covers)

radio systems in all five continents is available to prospective customers.

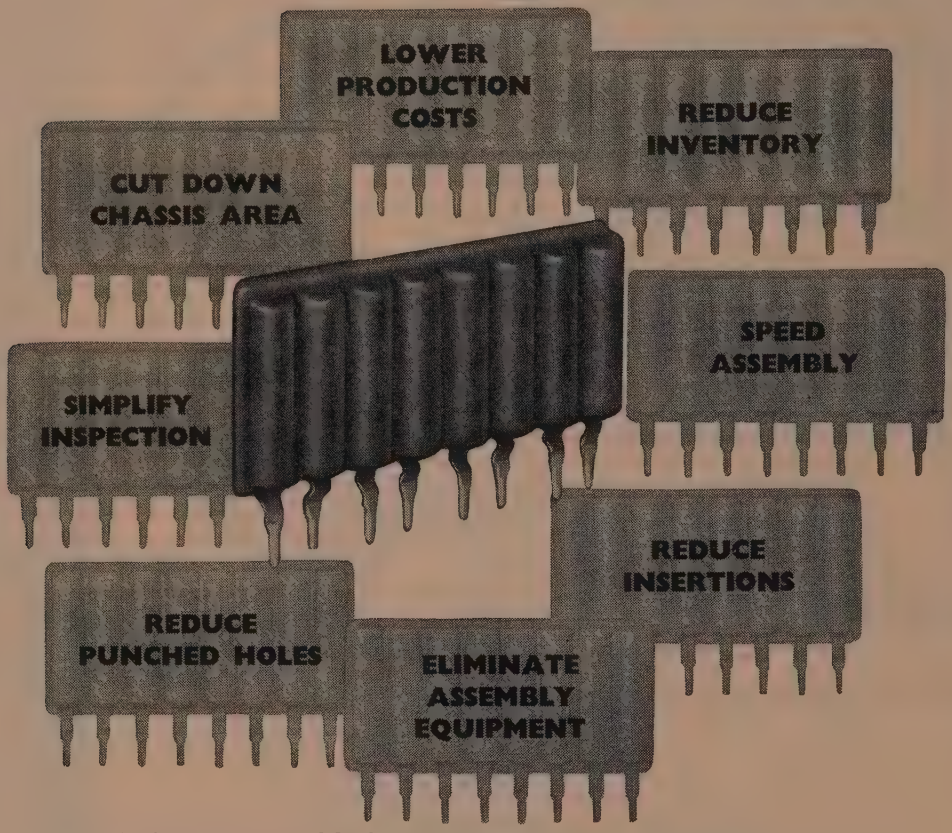
ATIONS

ONE, RADIO AND TELEVISION WORKS, COVENTRY

G.E.C.

G.E.C 24

Pre-assembled Components

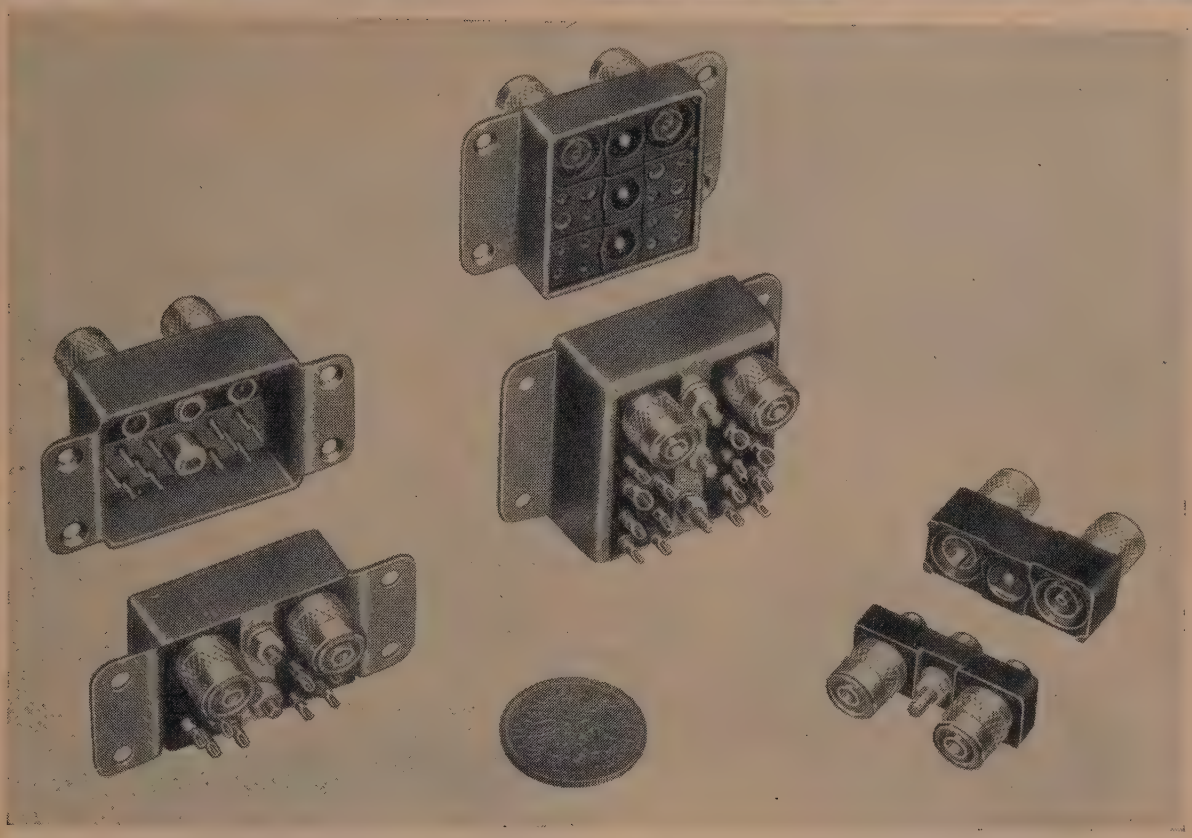


The Erie pre-assembled component system, known as "Pac", is a thin vertical module containing standard resistors and capacitors, of proven quality, mounted to a printed wiring board in a stable mechanical assembly, tailored to the requirements of the individual customer.

Besides offering considerable savings in itself, "Pac" paves the way for fully three-dimensional assembly, thus enabling the designer to take the fullest possible advantage of the potential savings in space attendant upon the introduction of shorter, wide-angle, tubes for television.

ERIE* *Resistor Ltd*

★ Registered Trade Mark



Actual size

The 'Domino' with a difference

"BELLING-LEE"
MINIATURE COAXIAL UNITORS
L1391/P and S

These Coaxial Unitors are designed to be used in conjunction with our present series of 'Domino' Unitors, and may be mounted with them in two- or three-way shrouds.

Each Unitor carries two miniature coaxial connectors which are of similar design to the plug and free socket of the L.1417 series and can be interconnected with them.

The brass shrouds (L.1404, L.1405 and L.1406) for mounting one, two or three Unitors, are each available for either flush or surface mounting.

Each Unitor is polarised by means of a flat on the locating spigot, so that the two coaxial ways cannot be accidentally interchanged.

ELECTRICAL DATA

D.C. breakdown voltage: Pressure:

Atmospheric	70 mm. of mercury	10 mm. of mercury	2 mm. of mercury
3750 V	1000 V	500 V	460 V

These figures relate to the voltage between the two-way shroud (L.1405) and one outer conductor. The Unitor was not loaded with cable.

MECHANICAL DATA

Weight: 1.04 oz. (29.5 g.)

Materials: Unitor body: Black Phenolic material X17163
Outer conductor: Aluminium alloy
Inner conductor: Gold-plated brass
Dielectric: P.T.F.E.
Circlip: Nylon

Maximum cable size over outer conductor: $\frac{1}{8}$" (3 mm.)

Withdrawal force: 4.5 lb. (2 kg.)

Most 'Belling-Lee' products are covered by patents or registered designs, or applications therefor.

BELLING & LEE LTD
GREAT CAMBRIDGE ROAD, ENFIELD, MIDDX., ENGLAND
Telephone: Enfield 3322 · Telegrams: Radiobel, Enfield

TERMINALS · PLUGS & SOCKETS · THERMAL DEVICES
FUSES · INTERFERENCE FILTERS · RECEIVING AERIALS

Marconi

V.H.F. SIGNAL GENERATOR TF 1064
and TRANSMITTER/RECEIVER OUTPUT TEST SET TF 1065

These two compact, portable units together provide full testing facilities for f.m. and a.m. mobile radio telephone sets. The range of tests which can be made is extremely comprehensive and for receivers includes sensitivity, bandwidth, signal-to-noise ratio, image rejection, discriminator symmetry, and a.f. power output. For transmitters, p.a. tuning, r.f. power output, and deviation can all be checked. If you would like further details of these versatile instruments, please write for Leaflets K138.



AM & FM SIGNAL GENERATORS · AUDIO & VIDEO OSCILLATORS
FREQUENCY METERS · VOLTMETERS · POWER METERS
DISTORTION METERS · FIELD STRENGTH METERS
TRANSMISSION MONITORS · DEVIATION METERS
OSCILLOSCOPES, SPECTRUM & RESPONSE ANALYSERS
Q METERS & BRIDGES

MARCONI INSTRUMENTS LTD · ST. ALBANS · HERTFORDSHIRE · TELEPHONE ST. ALBANS 56161
London and the South: Marconi House, Strand, London, W.C.2. Tel: COVent Garden 1234. *Midlands:* Marconi House, 24 The Parade, Leamington Spa. Tel: 1408
North: 23/25 Station Square, Harrogate. Tel: 67455.

TC 138

Printed Circuit Counter Panels



A complete range of transistorized counter panels of common size, fixing method and electrical connexion, designed to provide a flexible unit system whereby any special requirements in the counting or data processing fields can be quickly built up.

A fully illustrated brochure giving complete performance and specification figures for every panel in the range is available on request.

50kc/s Scaler
1Mc/s Scaler
Input Amplifier
Gate Unit
10kc/s Oscillator
1Mc/s Oscillator
Power Unit
50kc/s Read-out Scaler
1Mc/s Read-out Scaler
4 Channel Output Unit
Read-out Unit
Meter Display Unit
Lamp Display Unit
Numerical Indicator Tube
Shift Register Stage
Shift Register Driver

CINTEL

RANK CINTEL LIMITED
Worsley Bridge Road · London · SE 26
HITher Green 4600

Sales and Servicing Agents: Atkins, Robertson & Whiteford Ltd. Industrial Estate, Thornliebank, Glasgow; McKellen Automation Ltd., 122 Seymour Grove, Old Trafford, Manchester, 16; Hawnt & Co. Ltd., 59 Moor St. Birmingham, 4.

Semiconductors Limited

Applications Laboratory

making new equipment possible—existing equipment better . . .



The Semiconductors Limited Transistor Applications Laboratory, with greatly improved facilities, is now installed in a new building adjacent to the Swindon factory. This will considerably increase its capacity to assist manufacturers in the correct selection and use of Semiconductors transistors. You are invited to take advantage of this service, either directly or through your Semiconductors representative.

In addition to providing direct assistance to manufacturers, the Applications Laboratory is continuously investigating new approaches to existing problems.

For immediate assistance on your transistor problems telephone the Applications Laboratory SWINDON 6421.

transistors

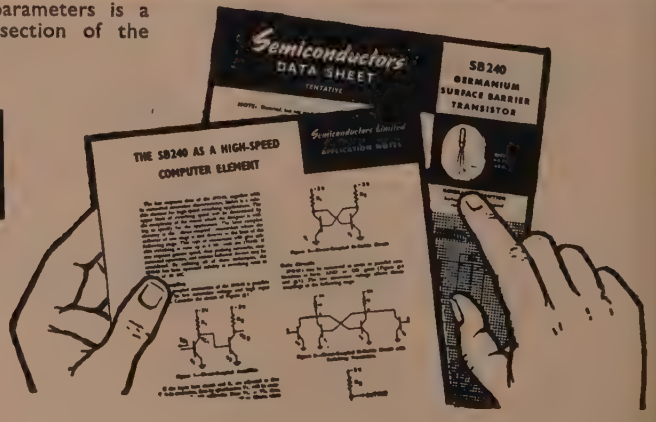


The evaluation of transistor high-frequency parameters is a continuing study in this specially-equipped section of the Transistor Applications Laboratory.

Semiconductors Limited

CHENEY MANOR · SWINDON · WILTSHIRE

Telephone: Swindon 6421/4 Telegrams: Semicon, Swindon

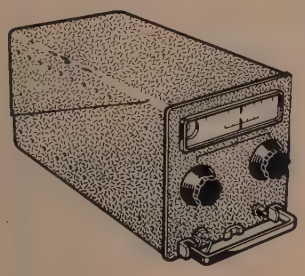


* *NEW APPLICATION NOTES AND DATA SHEETS ARE CONSTANTLY BEING ISSUED*
ARE YOU ON THE MAILING LIST?

SC.12

A.W.A ELECTRONICS

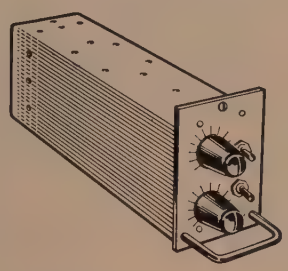
New concepts in electronics have been developed at AWA, as a result of experience with missile systems. Now they have a wider application. Here are some of the new AWA devices now available to industry.



U.H.F. WIDEBAND RECEIVER

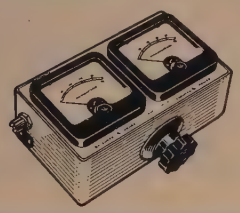
Basic arrangement consists of R.F. amplifier, mixer, local oscillator, I.F. amplifier (A.G.C. controlled), cathode follower output stage. Tuning indicator (EM 34) is also fitted to receiver. The standard forms: one for airborne racking with special separate power supply unit, the other on larger chassis including power supply unit (conventional 19″ front panel). *Standard specification: 420-470 M/cs frequency range; 4 M/cs overall bandwidth, approximately 10 db noise factor; approximately 70 ohms input impedance. 200-250 V and 50-60 c/s input supply. Input is unbalanced, output is via low impedance (cathode follower) stage.*

TRANSISTOR GALVANOMETER AMPLIFIER

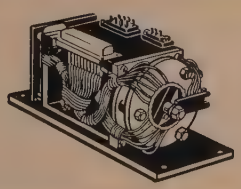


This Amplifier has been designed to drive viscous damped recording galvanometers which normally have a resistance of 50 ohms and a working range of DC to 2 Kc/s in frequency. The amplifier has a switched attenuator at its input and will accept single ended or push pull signals from ± 1 Millivolt to ± 500 volts and will feed a maximum of ± 50 Milliamps to the galvanometer. There is also a range of ancillary units available for use with this Amplifier as part of a comprehensive instrumentation system. *Standard specification: Dimensions: 4¼ in. x 3¾ in. x 10 in.; Frequency response: Flat from DC to 2 Kc/s, 5% down at 3 Kc/s, 3db down at 6 Kc/s; Noise Level: Less than 10 Microvolts; Input impedance: 40,000 ohms on range 5, 110,000 ohms all other ranges; Gain: Maximum 7.5 Milliamps/Millivolt, minimum 0.04 Milliamps/Volt; Power requirements: ± 6 Volts DC 220 Milliamps each line.*

DIRECTIONAL COUPLER



Of the 'Loop' type, suitable for measurements of RF power and Standing Wave Ratio in coaxial cables. Directional properties are largely unaffected by frequency changes, so coupler may be used to help obtain optimum termination of a 52 ohm coaxial system up to 600 M/cs. *Standard specification: Size 7″ x 4″ x 2½″; weighs 4 lbs. 3 ozs.; Power Measurement Range is Low range 1 w.cw.max. High range 5 w.cw.max.; less than 1% attenuation; better than 2% accuracy at frequency of calibration.*



ROTARY SWITCH FOR TELEMETRY

Based on a conception of British Ministry of Supply's Research and Development Establishment, gives facilities previously unobtainable from mechanical sampling devices. The Standard Model enables two 24 channel banks to be sampled at speeds up to 200 r.p.s.

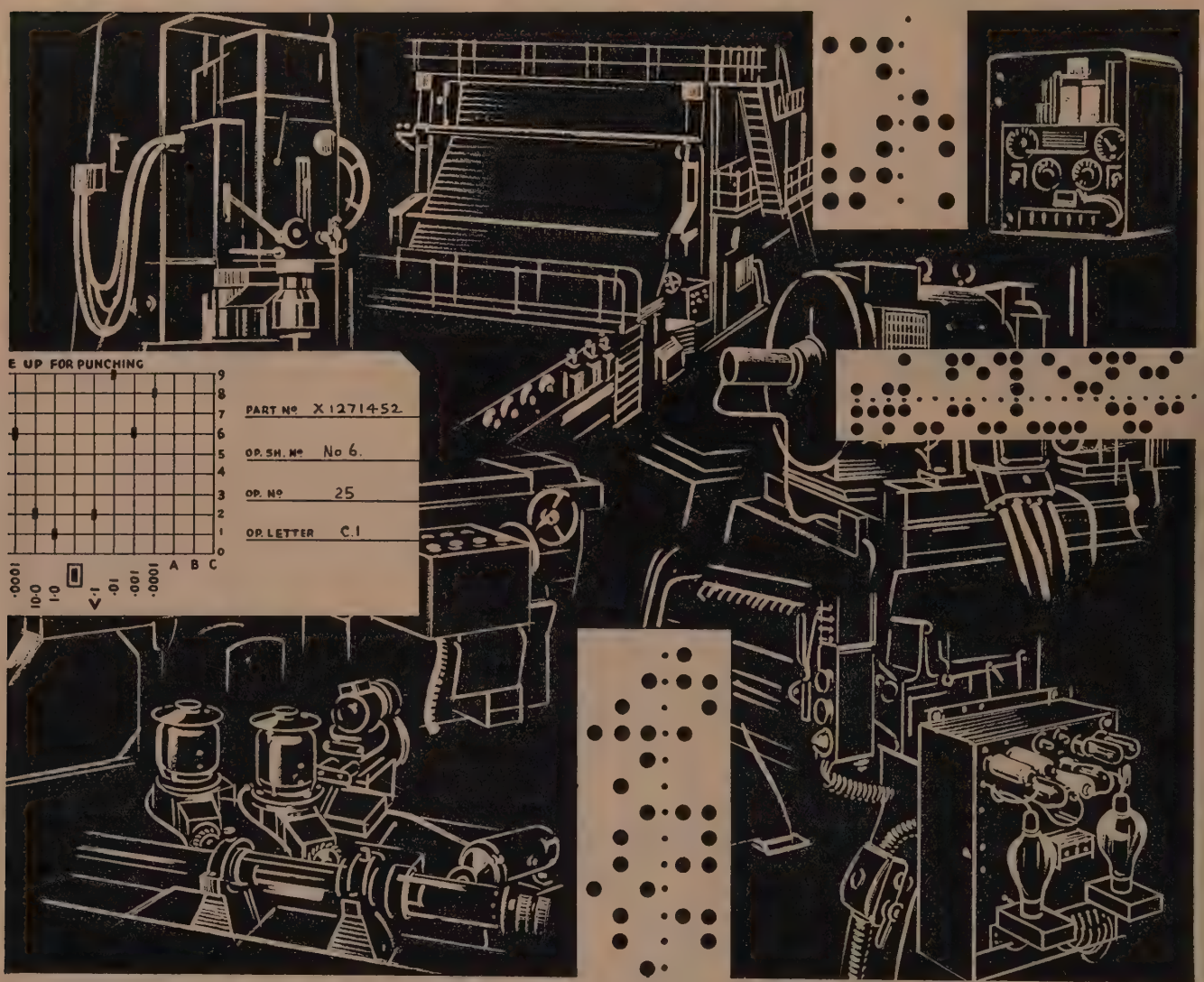
All devices are adaptable to suit customers' own requirements. For further information consult:

COMMERCIAL ELECTRONICS DEPT.

SIR W. G. ARMSTRONG WHITWORTH AIRCRAFT LTD., Baginton, Coventry.

MEMBER OF THE HAWKER SIDDELEY GROUP

Extending the applications of ELECTRONICS



through BTH Experience

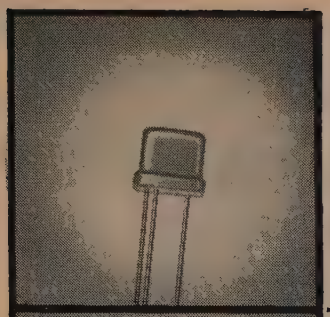
Machine-tool control
Motor speed control
Magnetic amplifiers
Vibration measuring equipment
High-frequency induction heating equipment
Infra-red relays and pyrometers
Timers and relays
Resistance welding control
Voltage regulators
Dynamic balancing equipment
Communication equipment
Photo-electric relays
Leak detectors

Write for booklet, 'BTH Electronics in Industry', to

BTH INDUSTRIAL ELECTRONICS SALES
LEICESTER, ENGLAND

THE BRITISH THOMSON-HOUSTON COMPANY LIMITED — an A.E.I. Company

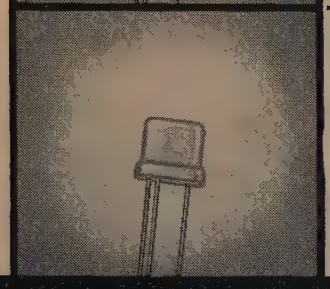
A5329



RF TYPES

		Min. f_α
6 Volts	V6/R8	8 Mcs.
	V6/R4	4 Mcs.
	V6/R2	2 Mcs.

Excellent general purpose R.F. transistors which are finding increasing uses in switching circuits. Have useful current gains at high pulse currents and can be selected for 15 and 30 Volt operation. Maximum dissipation now 75 mW; 125 mW on a heat sink. Clips supplied free.



AUDIO TYPES

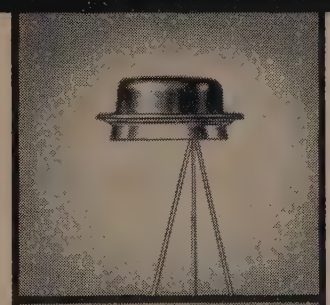
		Min. Beta
10 Volts	V10/50B	50
	V10/30A	30
	V10/15A	15

An ideal transistor for audio stages of portable receivers. Maximum dissipation now 125 mW; on a heat sink 200 mW. Clips supplied free. Selected types for 30 Volt operation can be supplied to order.

Newmarket GOLTOP* Transistors

NOW 4 RANGES IN PRODUCTION

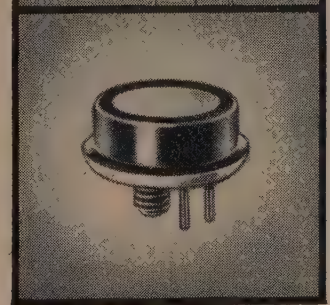
ALL TRANSISTORS ACTUAL SIZE



INTERMEDIATE POWER

Volts		Min. Beta 20
15	V15/20IP	
30	V30/20IP	

With maximum power dissipation of 2 watts, this transistor fills the gap between the audio types and the 10 watt power types in both power output and frequency range. Excellent for low-power transmitters at frequencies up to 500 kc/s. Sealed by cold welding and supplied with special clip mounting, eliminating large fixing studs. 60-Volt types can be selected to order.



POWER

	Minimum Beta		
Volts	10	20	30
15	V15/10P	V15/20P	V15/30P
30	V30/10P	V30/20P	V30/30P
60	V60/10P	V60/20P	V60/30P

The widest range of power transistors available with the closest tolerances. Extremely robust, reliable and stable. High heat dissipation—up to 20 watts per pair used in Class B with suitable heat sink. Up to 80 watts per pair obtainable in Inverter applications. Matched pairs can be supplied.

Leading the way With four years manufacturing experience behind us we are still expanding our production, still increasing our efficiency by new techniques in mechanisation. This policy ensures that the latest developments in transistors will always be offered by Newmarket at extremely competitive prices. The special requirements of every customer receive individual attention.

** Identified by the gold top.*

NEWMARKET TRANSISTORS

All enquiries to:

Newmarket
Transistor Co Ltd

Exning Road, Newmarket, Suffolk
Telephone: Newmarket 3381-4 Cables: Semicon Newmarket

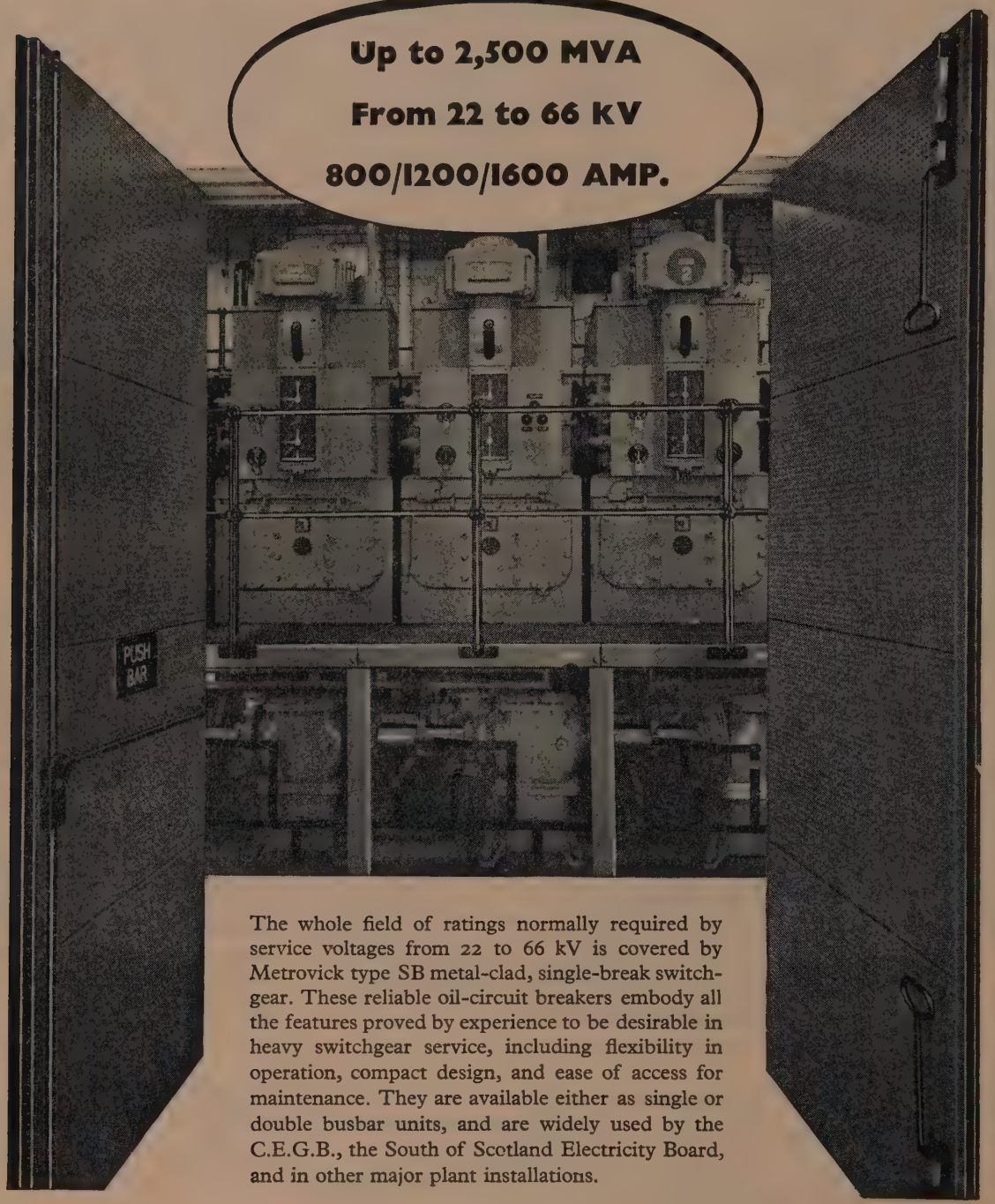
Single-break Switchgear

FOR INDOOR SERVICE

Up to 2,500 MVA

From 22 to 66 kV

800/1200/1600 AMP.



The whole field of ratings normally required by service voltages from 22 to 66 kV is covered by Metrovick type SB metal-clad, single-break switchgear. These reliable oil-circuit breakers embody all the features proved by experience to be desirable in heavy switchgear service, including flexibility in operation, compact design, and ease of access for maintenance. They are available either as single or double busbar units, and are widely used by the C.E.G.B., the South of Scotland Electricity Board, and in other major plant installations.

For further details please write for descriptive leaflet 292/19–1

METROPOLITAN-VICKERS
ELECTRICAL CO LTD · TRAFFORD PARK · MANCHESTER, 17

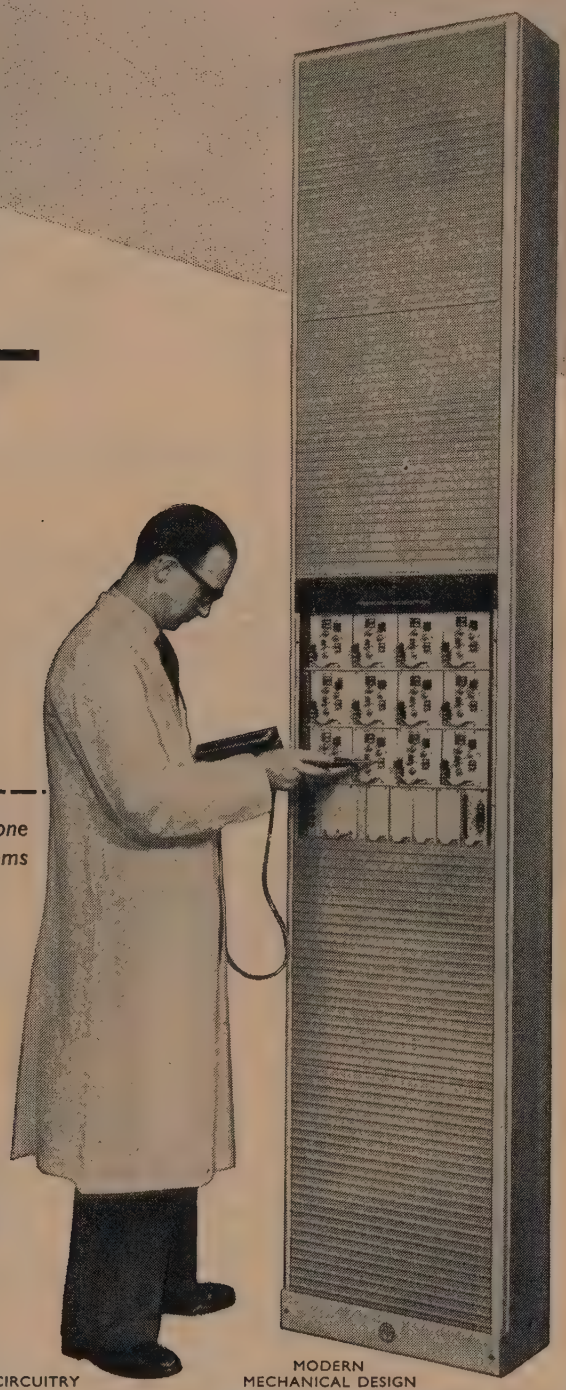
An A.E.I. Company

LEADING SWITCHGEAR PROGRESS

F/A801

A.T.E introduces–

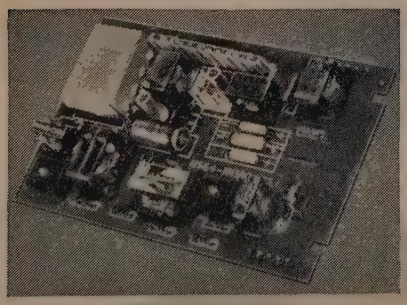
Transmission Equipment type C.M.



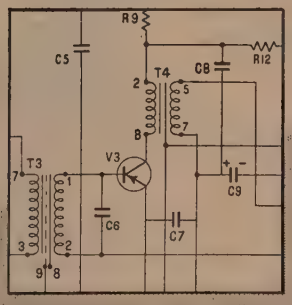
Illustrated is a rackside of telephone channelling equipment for cable or radio systems

- ✻ **Fully transistorized**
- ✻ **C.C.I.T.T. performance**
- ✻ **For existing or new installations**
- ✻ **Modern components, methods and design**
- ✻ **With inbuilt outband signalling**

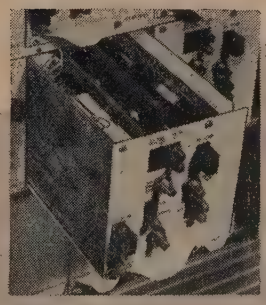
MODERN COMPONENTS



MODERN CIRCUITRY



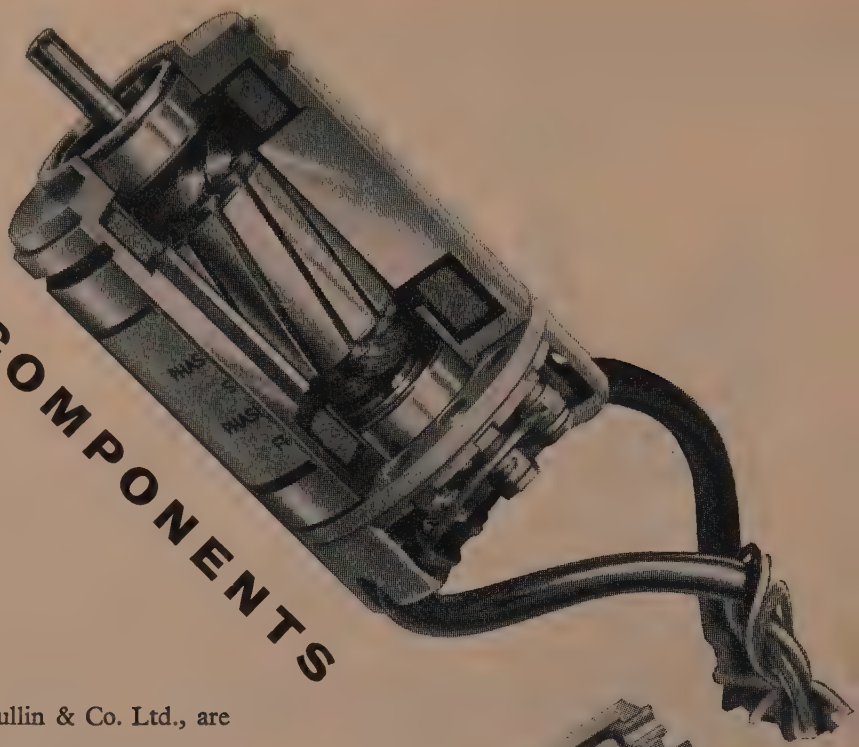
MODERN MECHANICAL DESIGN



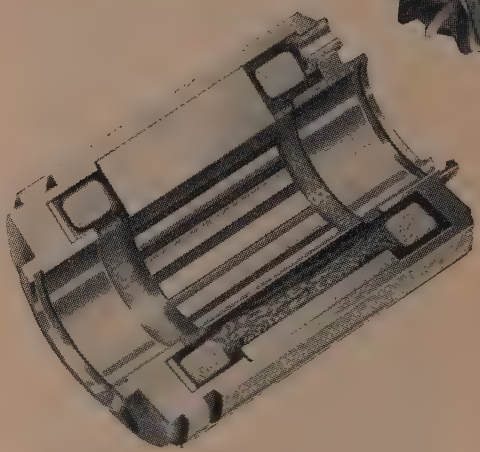
AUTOMATIC TELEPHONE AND ELECTRIC CO. LTD.
STROWGER HOUSE, ARUNDEL ST., LONDON, W.C.2

AT/8821

Araldite IN SERVOCOMPONENTS



The Synchro units shown, made by R. B. Pullin & Co. Ltd., are sectioned to show how the stators are integrally cast in Araldite to provide maximum protection against the effects of extremes of temperature, humidity and vibration. The excellent machining properties of Araldite make possible a straight-through bore technique, which eliminates errors in alignment and also permits the use of the smallest possible air gap between rotor and stator. High insulation and dielectric strength, remarkable adhesion to metals, and negligible shrinkage on curing make Araldite eminently suitable for use in the construction of precision electrical equipment.



Araldite epoxy resins are used—

- **for casting high grade solid electrical insulation**
- **for impregnating, potting or sealing electrical windings and components**
- **for producing glass fibre laminates**
- **for producing patterns, models, jigs and tools**
- **as fillers for sheet metal work**
- **as protective coatings for metal, wood and ceramic surfaces**
- **for bonding metals, ceramics, etc.**



This photograph shows an A.E.W. electric oven, capable of maintaining temperatures within close limits, as used by R. B. Pullin & Co. Ltd. for curing the Araldite-filled stators.

See us on Stand T8 at the ASEE Exhibition

Araldite *epoxy resins*

Araldite is a registered trade name

CIBA (A.R.L.) LIMITED

Duxford, Cambridge Telephone: Sawston 2121

AP 163



Marconi Crystals for stability and precision

The experience gained in manufacturing quartz crystals to the stringent requirements of our own apparatus and of Service equipment enables us to offer a comprehensive range of crystals covering the frequency band 1.0 Kc/s to 62 mc/s at extremely competitive prices.

Years of intensive research and development work guarantee the reliability and quality of this Marconi product.

MARCONI **CRYSTALS FOR ELECTRONICS**

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX, ENGLAND

M 5

Audio Power Valves

TY7-6000A



TY3-250



for
Relay Services
Public Address Systems
Vibration Equipment

Mullard audio power valves are available both for new equipment designs and maintenance. Full data for these power valves and details of xenon and mercury vapour rectifiers for associated power supplies are readily obtainable from the address below.

AUDIO POWER AMPLIFIER VALVES

	Type No.	pa max. (watts)	Va max. (kilovolts)	Ik max. (amps)	Power Output (2 valves) (kW)
For Maintenance	*Triodes*				
	MZ2–200	275	2.4	0.4	1.2
	MY3–275	275	3.0	0.5	1.3
For New Equipment	TY2–125	135	2.5	0.25	0.7
	TY3–250	250	3.0	0.48	1.3
	TY4–500	450	4.0	0.7	2.4
	TY7–6000A	6000	7.2	2.8	20
	Tetrodes				
	QY3–65	65	3.0	0.22	0.27
	QY3–125	125	3.0	0.32	0.55
	QY4–250	250	4.0	0.45	1.2

POWER RECTIFIER VALVES

	Type No.	PIV max. (kilovolts)	Max d.c. output current (amps)	Typical heating up time (secs)
Mercury Vapour	RG1–240A	6.5	0.25	60
	RG3–250A/866A	10	0.25	60
	RG3–1250	13	1.25	60
Xenon	RR3–250/ 3B28	10	0.25	10
	RR3–1250/ 4B32	10	1.25	30

Mullard

GOVERNMENT AND INDUSTRIAL VALVE DIVISION

Mullard

Mullard Limited · Mullard House
Torrington Place · London · WC1 · Tel: LANgham 6633

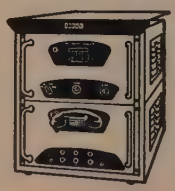
MVT 357b

ENGINEERING by RACAL

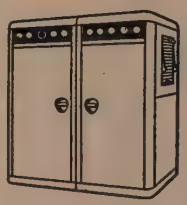
RADIO COMMUNICATIONS



RACAL ENGINEERING LTD., one of the more recent arrivals in the field of radio communications, have already made a name for original design and high quality. Their RA. 17 HF Communications Receiver has established a world-wide reputation and is being adopted in increasing numbers by the top communications authorities.

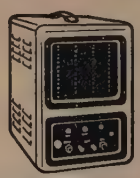


RACAL are ready to supply the increasing demand for Single Sideband radio links needed in modern communication systems. The TRA. 55 Transmitter/Receiver makes available even to small scale users all the advantages of SSB with remarkable simplicity of operation.



RACAL have an active radio development programme whose results are to be seen in a growing range of products precisely suited to the needs of today and tomorrow. Items now entering production include a comprehensive series of transmitters up to 5 kW, receivers and ancillary items.

DIGITAL INSTRUMENTATION



RACAL INSTRUMENT DIVISION offer a wide selection of valve operated Digital Counting Instruments assembled from plug-in units. Starting with the portable SA. 20 equipment, the range includes delay generators, chronometers and frequency meters supplemented by suitable transsducers for the majority of applications.



RACAL are well to the fore in the utilization of transistors for digital instrumentation. New standards of reliability and flexibility are achieved by the transistor-ferrite core plug-in cards employed in the construction of their latest range of instruments.



RACAL specialise in the design and manufacture of instrumentation systems employing standard equipment on a building-block basis. Their engineers, who are always available to discuss the problems and requirements of the factory and the laboratory, will be glad to provide unbiased advice and submit proposals.

These products and activities are sure to be useful to you. Why not write for full details now?

RACAL

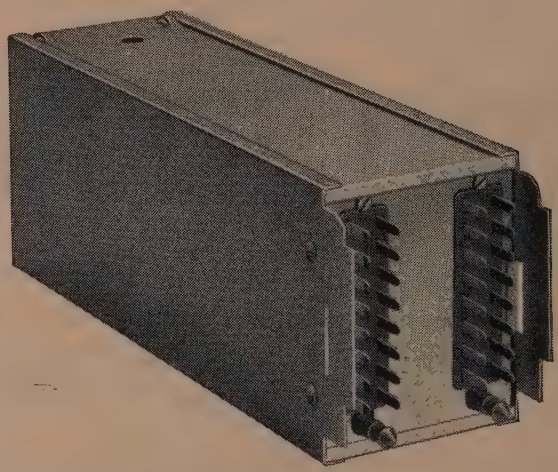
RACAL ENGINEERING LIMITED

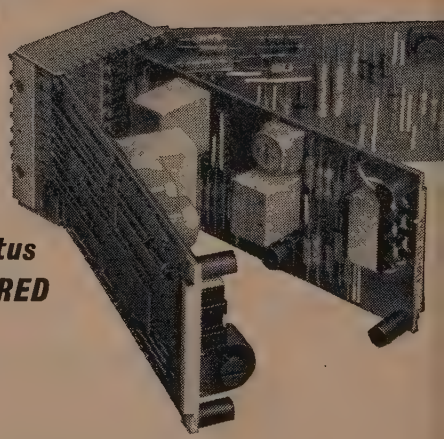
Western Road, Bracknell, Berkshire. Telephone: Bracknell 941 Telegrams/Cables: RACAL BRACKNELL BERKS
N. ENGLAND AGENT: Farnell Instruments, Ltd., Wetherby Industrial Estate, York Road, Wetherby, Yorks. Tel: Wetherby 2544
SCOTTISH AGENT: A. R. Bolton & Co., Ltd., 3A St. Vincent Street, Edinburgh. Tel: Edinburgh 32035
OVERSEAS: Agents operate in the majority of territories throughout the world

STC



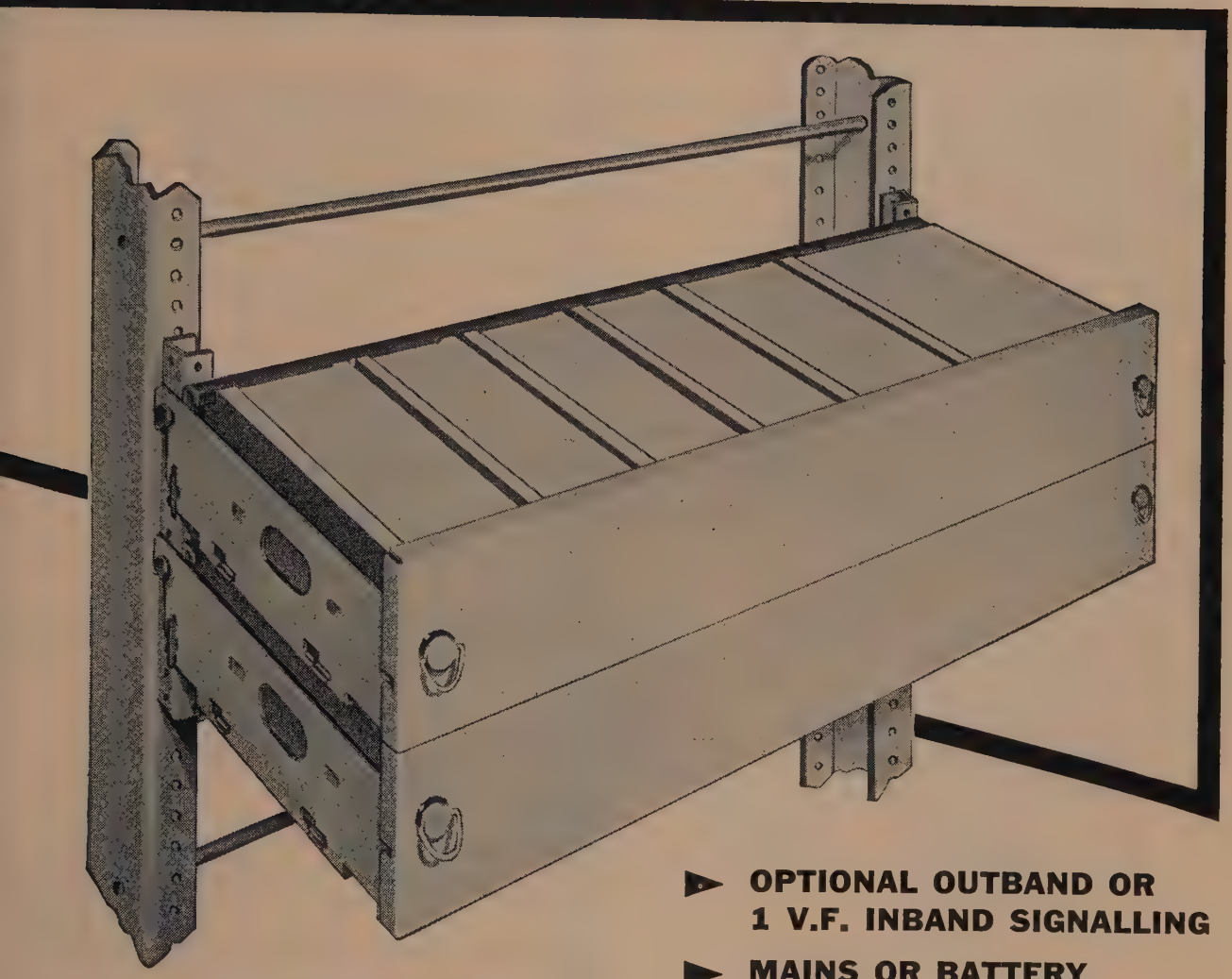
The new 12-circuit group sub-rack—a com
wired and tested mechanical unit. A no
advance in the design of carrier telep
equipment.





Standard Transmission Equipment Practice No. 5 uses plug-in apparatus units with STRIP-WIRED boards giving full access to the components.

L-TRANSISTOR CHANNELLING EQUIPMENT TYPE S.T.E.P. 5



- DOUBLED EQUIPMENT DENSITY
- 144 WIDE BAND CIRCUITS ON ONE STATION FRAME
- LOW POWER CONSUMPTION
- OPTIONAL OUTBAND OR 1 V.F. INBAND SIGNALLING
- MAINS OR BATTERY OPERATION
- LESS ROOM HEATING
- TEST POINTS EASILY ACCESSIBLE

Standard Telephones and Cables Limited

Registered Office: Connaught House, Aldwych, London, W.C.2

TRANSMISSION DIVISION: NORTH WOOLWICH · LONDON E.16

CATHODEON CRYSTALS LIMITED

H. F. BAND PASS CRYSTAL FILTERS

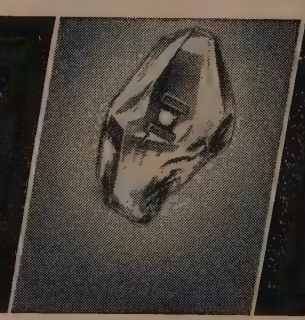
—on a centre frequency of 10.7 Mc/s possessing 80 dB of attenuation outside the pass-band are now available to designers of telecommunication equipment requiring the most stringent selectivities. They are eminently suitable for all AM/FM receivers operating in the V.H.F., U.H.F., bands, and require only 2″ x 1.5/16th″ of chassis space. They give substantially flat pass-band characteristics with almost vertical cut-offs. Several band-widths are available and all possess minimum attenuations of 80 dBs. Relatively low insertion loss values renders their use between the frequency changer and the I.F. amplifier most suitable and the characteristics entirely eliminate the need for I.F. transformers and permit broad band amplifier design techniques with only a single frequency conversion.

OPERATING CHARACTERISTICS (Provisional Data)

	Filter Type		
	BP 50	BP 25	BP 12
Centre Frequency	10.7 Mc/s	10.7 Mc/s	10.7 Mc/s
Centre Freq: Stability	± 0.0012%	± 0.0012%	± 0.0012%
Bandwidth at — 6 dB	25 kc/s	12.5 kc/s	6 kc/s
Bandwidth at — 60 dB	< 45 kc/s	< 22 kc/s	< 11 kc/s
Min. Attenuation outside pass band	80 dB	80 dB	80 dB
Power insertion loss in pass band	< 6 dB	< 6 dB	< 6 dB
Terminating Impedance	1600Ω	1000Ω	1000Ω
Operating Temp: Range	—40° to + 90°C	—40° to + 90°C	—40° to + 90°C
Weight	5½ oz.	5½ oz.	5½ oz.

Write or phone for full details including valve or transistor terminating circuits. (Patent application No. 37781/58.)

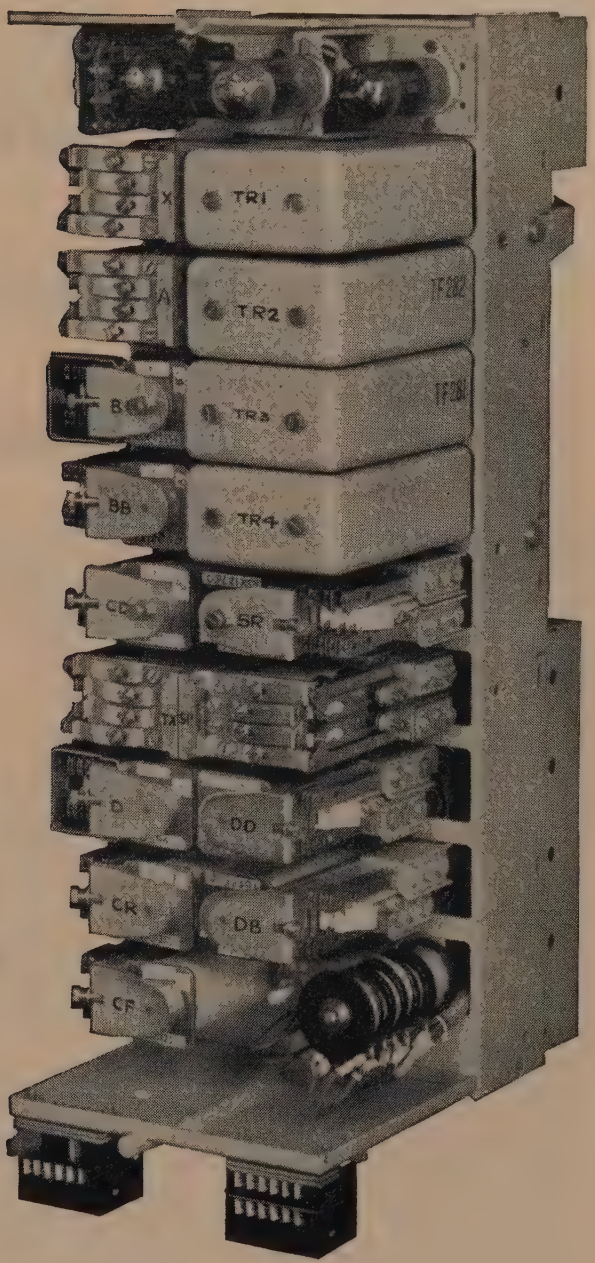
CATHODEON CRYSTALS LIMITED
LINTON CAMBRIDGESHIRE
TELEPHONE LINTON 501 (3 lines)



Economic trunk dialling

with 1VF

For many years Siemens Edison Swan has been one of the main world suppliers of Voice Frequency Signalling Equipment for large trunk networks. To meet the needs of local junction and minor trunk networks not requiring the more complex signalling facilities of two-tone systems, an advanced single voice frequency signalling and dialling equipment is also available at low cost.



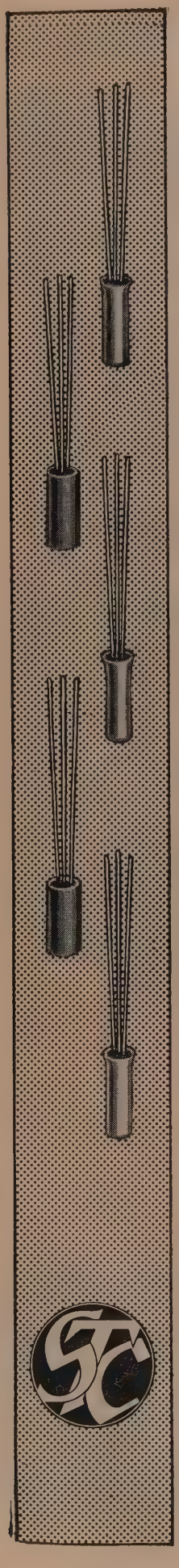
1VF Terminal Relay Set

The design of the system incorporates C.C.I.T.T. performance recommendations. All circuits in a network capable of transmitting speech may be adapted economically for through signalling and dialling. The following are some of the salient features of the equipment:—

- Can be applied to any speech channel
- Basically simple—does not complicate the line terminal or channelling equipment
- Can be arranged for unidirectional or bothway signalling and dialling
- Relay Set mounting in accordance with latest telephone exchange practice; can be installed with the telephone equipment and served from exchange power supplies
- Uses standard components of latest design throughout
- Requires negligible maintenance attention
- Suitable for the introduction of economic subscriber trunk dialling

SIEMENS EDISWAN

SIEMENS EDISON SWAN LIMITED
An A.E.I. Company
Telecommunications Division P.D.5
Woolwich, London, S.E.18
Telephone: Woolwich 2020



You can now specify S.T.C. Transistors for Entertainment and Industrial Applications

ENTERTAINMENT TYPES

(a) TS.7 TS.8 ... ...small signal amplifiers and frequency changers

(b) TS.9 TS.17 ... ...higher power amplifiers up to several hundred milliwatts

(c) TS.13 TS.14 ... ...low frequency applications

INDUSTRIAL TYPES

(a) TK.23A ... ...low frequency telephone and telegraph carrier systems

(b) TK.20B ... ...switching circuits (computers)

TK.25B ... ...switching circuits 8 Mc/s and above

(c) TK.21B TK.24B

TK.26B TK.27B ...switching circuits and/or small signal amplification

(d) TK.40A ... ...amplifier and oscillator. Audio frequencies for power of several hundred milliwatts

First in Europe to produce transistors, S.T.C. are now making available to industry a range previously only used in their own equipment. Their long experience of components manufacture coupled with the latest production techniques has resulted in yet another high-quality product.

Send for literature on the available types.

Standard Telephones and Cables Limited

Registered Office: Connaught House, Aldwych, London, W.C.2

TRANSISTOR DIVISION: FOOTSCRAY · SIDCUP · KENT

THEY SET A STANDARD

OVER 1500 years ago the Romans built the roads whose names are still familiar to us—Fosse Way, Watling Street, Icknield Way, Ermine Street and many others. Often following tracks old before the Romans came, these broad, straight roads provided vital communication for the legions, and for the administrators and merchants who followed in their steps. In road building, as in many other fields, the Romans set a standard which few have equalled since.

In cable making, too, standards are of vital importance. For over 100 years members of the Cable Makers Association have been concerned in all major advances in cable making. Together they spend over one million pounds a year on research and development. The knowledge gained is available to all members. This co-operation has contributed largely to the world-wide prestige that C.M.A. cables enjoy, and it has put Britain at the head of the world cable exporters.

Technical information and advice is freely available from any C.M.A. member.

Roman legionaries on the march.



MEMBERS OF THE C.M.A.

British Insulated Callender's Cables Ltd · Connollys (Blackley) Ltd · Enfield Cables Ltd · W. T. Glover & Co. Ltd · Greengate & Irwell Rubber Co. Ltd. W. T. Henley's Telegraph Works Co. Ltd · Johnson & Phillips Ltd. The Liverpool Electric Cable Co. Ltd · Metropolitan Electric Cable & Construction Co. Ltd · Pirelli-General Cable Works Ltd. (The General Electric Co. Ltd.) · St. Helens Cable & Rubber Co. Ltd · Siemens Edison Swan Ltd · Standard Telephones & Cables Ltd · The Telegraph Construction & Maintenance Co. Ltd.

Insist on a cable with the C.M.A. label

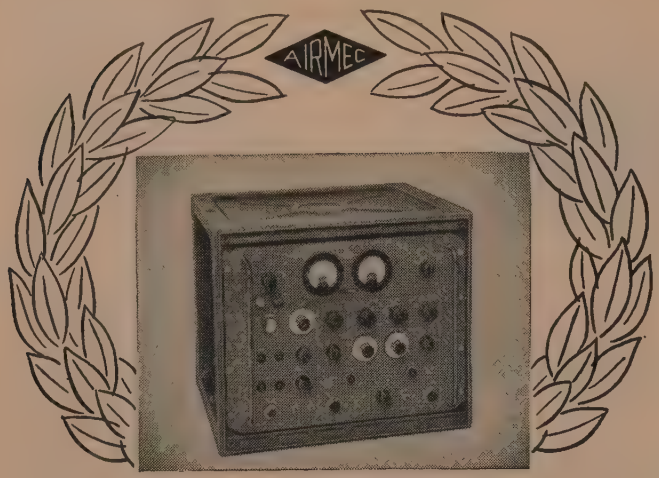


The Roman Warrior and the letters 'C.M.A.' are British Registered Certification Trade Marks.

CABLE MAKERS ASSOCIATION

ABLE MAKERS ASSOCIATION, 52-54 HIGH HOLBORN, LONDON, W.C.1. TELEPHONE: HOLBORN 7633

CMA 23



KLYSTRON POWER UNIT

TYPE 1856

Developed and manufactured for the
MINISTRY OF SUPPLY
in conjunction with the
ROYAL RADAR ESTABLISHMENT

Now offered for general use, the Klystron Power Unit enables low voltage Klystron oscillators to be operated with an exceptionally high degree of frequency and power stability.

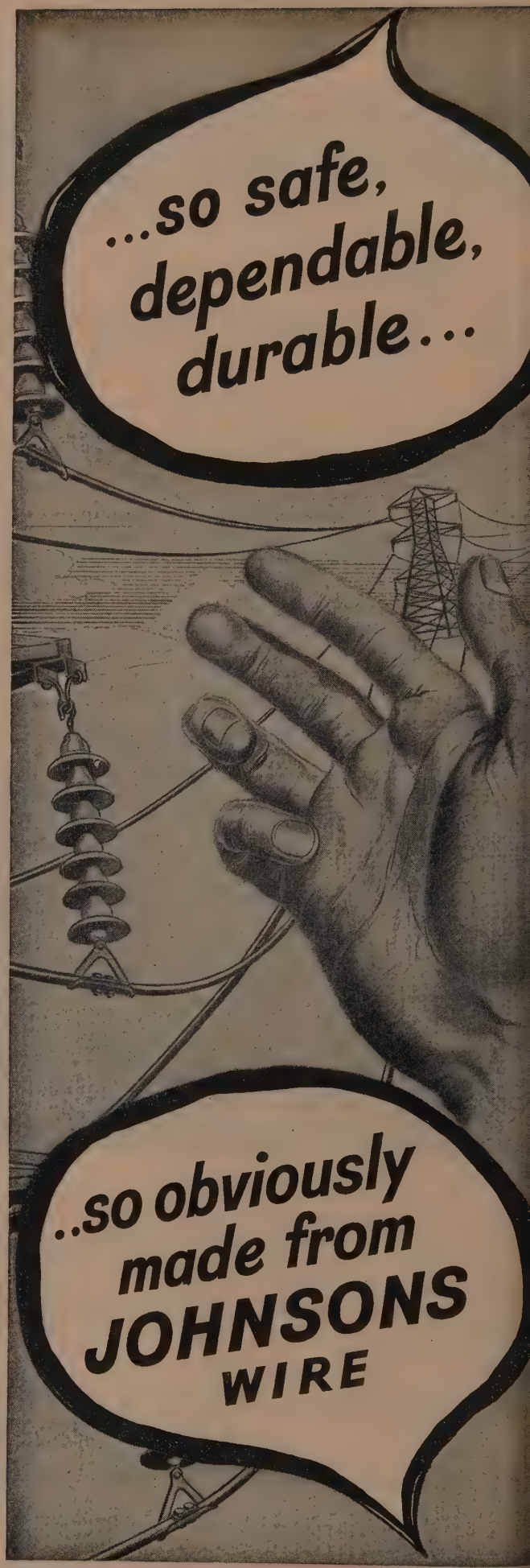
SPECIAL FEATURES

- **2 Power Supplies in one unit for operating Klystrons either singly or in pairs—separate controls for each supply.**
- **Modulation—square and saw-tooth wave provided by internal oscillators.**
- **Highly stable ten turn helical potentiometer for adjustment of the cathode output voltage.**

Send for leaflet 176

Airmec

AIRMEC LIMITED · HIGH WYCOMBE · BUCKS
TELEPHONE: HIGH WYCOMBE 2060-4



RICHARD JOHNSON & NEPHEW LIMITED, MANCHESTER 11 Tel. EAST 1

over 4 megawatts

Stabilised by SERVOMEX



VOLTAGE STABILISER A.C.7 (30 Amp)
(shown with wooden ends, an optional extra)
Price £143 *rack mounted. Delivery from stock.*

Servomex Voltage Stabilisers are constantly being improved without increase in prices

The installed capacity of Servomex A.C.7 voltage Stabilisers now runs to over 4 Megawatts. They have been widely specified by Government Research establishments and many of the premier industrial concerns and instrument makers throughout the world.

It is becoming increasingly common to stabilise the mains supply at the source, thereby optimising the performance of electronic apparatus and prolonging its life.

Write for Data Sheets and a copy of "Technical Notes on A.C. Stabilisers".

A.C. VOLTAGE STABILISERS: A.C.7 (30 Amp.) A.C.2 Mk. IIA (9 Amp.) A.C.2 Mk IIB (9 Amp.)

These three instruments use a variable transformer driven by a continuous servo system with velocity feedback. All three have been re-designed in our "Ruggedised" range; all valves, resistors and capacitors are derated strictly in accordance with current Inter-Service Requirements. There are no relays, thyratrons, or electrolytic capacitors. The whole apparatus is immensely strong, both mechanically and electrically. Ernest Turner moving iron meters are provided for voltage and current except on the A.C.2 Mk. IIA.

The following apply to both instruments

★ *Zero distortion* ★ *Accuracy 0.25%* ★ *Efficiency 95% or better* ★ *Unaffected by frequency changes* ★ *True r.m.s. stabilisation* ★ *Load Power factors zeros to unity lead or lag*

A.C.2 Mk. IIA and IIB
★ *0 to 9 amps*
★ *Range minus 17.5% to plus 8.75%*
★ *15 volts per second*

A.C. 7
★ *0 to 30 amps.*
★ *Range minus 20% to plus 10%*
★ *12 volts per second*

SERVOMEX CONTROLS LIMITED • CROWBOROUGH • SUSSEX • TELEPHONE: CROWBOROUGH 1247

PHILIPS

In Science and Industry alike . . .

among technicians, manufacturers and those engaged in the sale of electrical products — as well as among the public at large, the Philips emblem is accepted throughout the World as a symbol of quality and dependability.

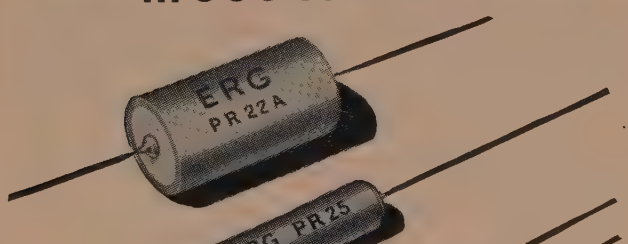
PHILIPS ELECTRICAL LTD

Century House · Shaftesbury Avenue · London · WC2

Radio & Television Receivers · Radiograms & Record Players · Gramophone Records · Tungsten, Fluorescent, Blended and Discharge Lamps & Lighting Equipment · 'Philishave' Electric Dry Shavers · 'Photoflux' Flashbulbs · High Frequency Heating Generators · X-ray Equipment for all purposes · Electro-Medical Apparatus · Heat Therapy Apparatus · Arc & Resistance Welding Plant and Electrodes · Electronic Measuring Instruments · Magnetic Filters · Battery Chargers and Rectifiers · Sound Amplifying Installations · Cinema Projectors · Tape Recorders · Health Lamps · Hearing Aids · Electrically Heated Blankets

The PRECISION WIRE-WOUND Resistor

—most in demand



Other products include Vitreous coated Silicone and Glass Bond Resistors, Transformers, Chokes & Interleaved Coils.

ERG

Trade Mark

ERG INDUSTRIAL CORPORATION LTD
18 MANCHESTER STREET · LONDON · W.1
Tel: WELbeck 8114/5

ADCOLA (Regd. Trade Mark) Soldering Instrument

ILLUSTRATED
PROTECTIVE SHIELD (CAT. No. 68)
⅛ in. BIT MODEL (CAT. No. 70)

Primarily developed for the TRANSISTOR & ELECTRONIC ERA.

Possessing the sharp heat essential for the quick jointing of Transistors, Resistors, etc., thereby avoiding damage to the equipment from heat transference



Cover all requirements for thorough solder jointing in all the fields of

TELECOMMUNICATION

Fully Insulated Elements.

Suited to daily use for bench line production

MANUFACTURED IN ALL VOLT RANGES

British and Foreign Pat. Reg. designs, etc.

For further information apply Head Office:

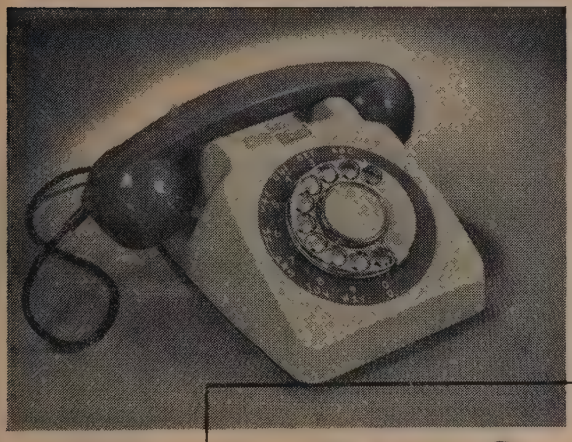
ADCOLA PRODUCTS LTD.
GAUDEN ROAD
CLAPHAM
HIGH STREET
LONDON, S.W.4

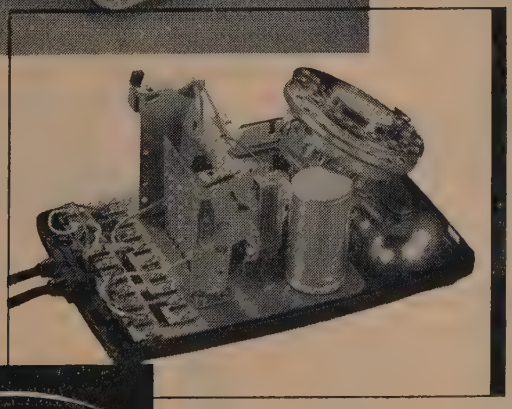
Tel.: MACaulay 4272 & 3101

The new British Post Office Telephone



but fully tropicalised





Anticipating the demand from many overseas countries for the new British Post Office telephone, Siemens Edison Swan have developed a model suitable for use in any climate. This instrument is in full tropical finish, ventilated and effectively sealed against dust and insects. Available in colours.

It can be supplied with or without automatic transmission regulator and an attenuator for the transmitter is also an optional feature. When the automatic regulator is not supplied, provision is made for adding it at any time.

SIEMENS EDISWAN

For full particulars write to:

SIEMENS EDISON SWAN LTD.

A.E.I. Company

Telecommunications Division P.D.7/4

Woolwich, London, S.E.18

Telephone: Woolwich 2020 Ext. 724

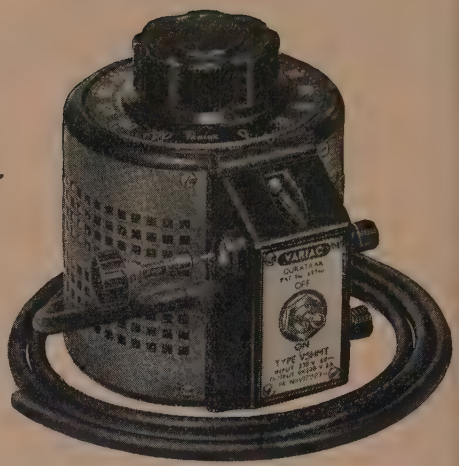
Variac with Duratrak*

for high-efficiency voltage control

VARIAC is the original continuously-adjustable auto-transformer, providing a smoothly variable output voltage from zero to 17% above line. Ultra-low-loss 'rolled' core and DURATRAK (patent) track surface give very high efficiency at all settings. A very wide range of models is available, from small units (e.g. type V-5, illustrated on right) for laboratory and instrument use to large ganged assemblies for three-phase power.

VARIACS can be supplied open or covered, as single units or ganged assemblies of two or more units, for manual operation or motor-driven. The range includes portable models, metalclad models, dual-output models, high-frequency types, and many 'specials'.

*DURATRAK is a rhodium-covered, silver-plated track surface, which inhibits oxidation of the brush track and greatly reduces contact resistance, giving longer life, increased overload and surge capacity and maximum economy in maintenance.



Type V-5HMTF, a small VARIAC for laboratory use. Output 0-270 V, continuously adjustable, from 230 V 50 c/s mains. Rated current 2 A. Provided with terminals, switch, fuse and 3-core lead.

For complete information on the entire VARIAC range, request **Catalogue 424-UK**

Claude Lyons Ltd.

VARIAC and DURATRAK are registered trademarks DURATRAK is protected by U.K. patent No. 693406

Only VARIAK has DURATRAK

76 OLD HALL ST LIVERPOOL 3 · Tel: CENTRAL 4641-2

VALLEY WORKS · HODDESDON · HERTS · Tel: HODDESDON 3007-8-9

CL/46/E2A

ZENITH

(REGD. TRADE-MARK)

Automatic

VOLTAGE REGULATORS

with Electronic Control

A large range of stabilisers is available for installations requiring a constant voltage within ±1% from a fluctuating up to ±10% or −15% to +5%.

Single and three phase models from 7 to 25 kVA per phase are detailed in our catalogue which will be sent on request.



The ZENITH ELECTRIC CO. Ltd.

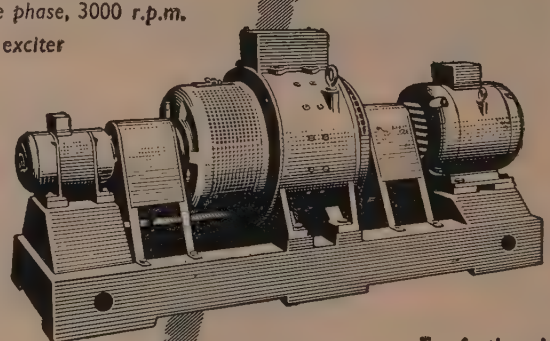
ZENITH WORKS, VILLIERS ROAD, WILLESDEN GREEN
LONDON, N.W.2

Telephone: WILlesden 6581–5 *Telegrams:* Voltaohm, Norphone, London

MANUFACTURERS OF ELECTRICAL EQUIPMENT
INCLUDING RADIO AND TELEVISION COMPONENTS

ELECTRICAL EQUIPMENT

Motor Alternator Set
12 kVA, 400 c.p.s.
single phase, 3000 r.p.m.
with exciter



For further details write to the manufacturers.

NEWTON DERBY

NEWTON BROS. (DERBY) LTD

ALFRETON ROAD DERBY

PHONE. DERBY 47676 (4 LINES) GRAMS· DYNAMO, DERBY
London Office: IMPERIAL BUILDINGS, 56 KINGSWAY W.C 2

NEW— from

Redifon

GR.310 1 watt H.F.

TRANSISTORISED

WALKIE - TALKIE

for telephony and CW



4 crystal-controlled spot frequencies anywhere in the range 2.5 to 8 Mc/s

LIGHT AND COMPACT
RUGGED AND TROPICALISED
NO TUNING
SIMPLE TO OPERATE

Brief specification

Frequency Range: 2.5 to 8 Mc/s.
Power Output: 1 watt.
Receiver Sensitivity: 3—6 microvolts.

Batteries: Self-contained chargeable lightweight silver-zinc cells. Special charging board available for plug-in block charging.

Consumption: 20 hours continuous operation with 5 to 1 duty cycle on one charge.

Full details on request

Radio Communications Division
REDIFON LIMITED, BROOMHILL ROAD, LONDON, S.W.18. **VANdyke 7281**

A Manufacturing Company in The Rediffusion Group

100R

Publications of

THE INSTITUTION OF ELECTRICAL ENGINEERS

Journal of The Institution—Monthly

Proceedings of The Institution

PART A (Power Engineering)—Alternate Months
PART B (Radio and Electronic Engineering—including Communication Engineering)—Alternate Months
PART C (Institution Monographs)—In collected form twice a year

Special Issues

VOL. 94 (1947) PART IIA (Convention on Automatic Regulators and Servomechanisms)
VOL. 94 (1947) PART IIIA (Convention on Radiocommunication)
VOL. 97 (1950) PART IA (Convention on Electric Railway Traction)
VOL. 99 (1952) PART IIIA (Convention on the British Contribution to Television)
VOL. 100 (1953) PART IIA (Symposium of Papers on Insulating Materials)
Heaviside Centenary Volume (1950)
Thermionic Valves: the First 50 years (1955)
VOL. 103 (1956) PART B SUPPLEMENTS 1-3 (Convention on Digital-Computer Techniques)
VOL. 103 (1956) PART A SUPPLEMENT 1 (Convention on Electrical Equipment for Aircraft)
VOL. 104 (1957) PART B SUPPLEMENT 4 (Symposium on the Transatlantic Telephone Cable)
VOL. 104 (1957) PART B SUPPLEMENTS 5-7 (Convention on Ferrites)
VOL. 105 (1958) PART B SUPPLEMENT 8 (Symposium on Long Distance Propagation above 30 Mc/s)
VOL. 105 (1958) PART B SUPPLEMENT 9 (Convention on Radio Aids to Aeronautical and Marine Navigation)
VOL. 105 (1958) PART B SUPPLEMENTS 10-12 (International Convention on Microwave Valves)

★

PROCEEDINGS - *Paper and Reprint Service*

PAPERS READ AT MEETINGS

Papers accepted for reading at Institution meetings and subsequent republication in the *Proceedings* are published individually without delay, free of charge. Titles are announced in the *Journal* of The Institution, and abstracts are published in *Science Abstracts*.

REPRINTS

After publication in the *Proceedings* all Papers are available as Reprints, price 2s. (post free). The Reprint contains the text of the Paper in its final form, together with the Discussion, if any. Those who obtain a copy of a Paper published individually—if they do not take the Part of the *Proceedings* in which it will be republished—are urged to apply in due course for a Reprint, as this is the final and correct version.

CONVENTION PAPERS

Papers accepted for presentation at a Convention or Symposium, and subsequent republication as a Supplement to the appropriate part of the *Proceedings*, are published shortly before the Convention, but are usually available only in sets. No Reprints are available.

MONOGRAPHS

Institution Monographs (on subjects of importance to a limited number of readers) are available separately, price 2s. (post free). Titles are announced in the *Journal* and abstracts are published in *Science Abstracts*. The Monographs are collected together and republished twice a year as Part C of the *Proceedings*.

An application for a Paper, Reprint or Monograph should quote the author's name and the serial number of the Paper or Monograph, and should be accompanied by a remittance where appropriate. For convenience in making payments, books of five vouchers, price 10s., can be supplied.

SCIENCE ABSTRACTS
Published monthly in two sections
SECTION A: Physics
SECTION B: Electrical Engineering

Prices of the above publications on application to the Secretary of The Institution, Savoy Place, W.C.2.

Permanent Magnets

Design Advisory Service

No. 12

Magnetic Attraction Applications—1

Advertisements in this series deal with general design considerations. If you require more specific information on the use of permanent magnets, please send your enquiry to the address below, mentioning the Design Advisory Service.

The earliest known effect of permanent magnets is their ability to attract ferrous objects.

The attraction or holding power of a magnet under ideal conditions can be calculated from the basic formula :—

$$\text{Force in dynes} = \frac{B^2A}{8\pi}$$

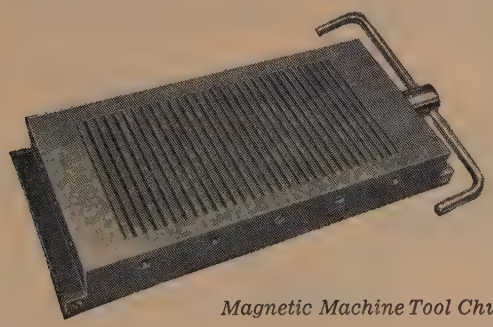
or

$$\text{Force in lb.} = \frac{B^2A}{11{,}263{,}000}$$

where A = *area of magnetic pole faces in cm²*
and B = *corresponding flux density in gauss.*

Under normal conditions joints, tolerances, misalignments and leakage flux rapidly reduce the theoretical pull; therefore the calculated value should only be used as a guide. One factor which must always be observed is that the magnet under the most severe conditions of open circuit should be of sufficient length to prevent self-demagnetisation —i.e. the maximum value of H should not exceed H_d.

This advertisement deals briefly with some of the many industrial applications using magnetic attraction, such as Machine Tool Chucks, Relays, Industrial Filtration, Door Catches, etc.



Magnetic Machine Tool Chuck

Machine Tool Chucks

One of the most useful applications of permanent magnets is in machine tool chucks where steel articles of a form extremely difficult to clamp, are held firmly in position for machining operations.

There are many types of magnetic chuck designed for various purposes but generally with relatively wide pole pitch spacing. The chuck illustrated has the advantage of small pole spacing and is particularly suitable for small or thin articles. It consists of thin 'Magnadur' blocks with mild steel pole plates assembled in sandwich form giving alternate poles 1/8″ apart.

The attractive force to iron and steel objects of not less than 1/16″ in thickness is approximately 130 lb./sq. in. The objects are released by moving the lower section of the chuck one pole pitch to short circuit or cancel the flux in the whole chuck.

Industrial Filtration Equipment

It is well known that one of the principal causes of wear in machinery is the presence of abrasive matter in the lubricating oil. A certain amount of the contamination to the oil can be prevented by careful design and dust covers but minute particles of steel can only be removed by the use of permanent magnets.

Relays and Thermostats

For current carrying relays and thermostats, contacts should open and close with a snap action and contact pressure must be sufficient to prevent chatter and arcing. It is in applications such as these that the attractive force of a magnet can be used to supply the necessary minimum contact pressure and also the desired degree of snap action.

Magnetic Fishing Tool

The tool shown is used for recovering broken rock drills or bits of iron or steel which accidentally get into deep boreholes. The one illustrated is of 14″ diameter, and uses a 'Ticonal' magnet capable of lifting over two tons. *Photograph by courtesy of D. F. J. Burns Co., Ltd.*



Magnetic Door Catches

Magnetic door catches can be designed to be extremely small, efficient and inexpensive. As an example, a 'Magnadur' magnet 0.89″ x 0.59″ x 0.18″, when fitted between mild steel pole plates, is capable of holding an armature with a force of between four and five pounds.

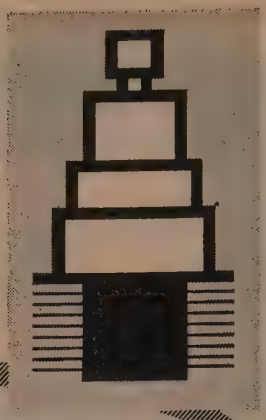
If you wish to receive reprints of this advertisement and others in this series write to the address below.

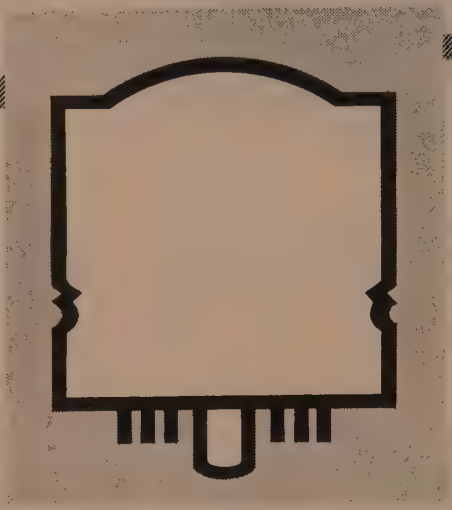
Mullard

'TICONAL' PERMANENT MAGNETS
'MAGNADUR' CERAMIC MAGNETS
FERROXCUBE MAGNETIC CORES

MULLARD LIMITED, COMPONENT DIVISION, MULLARD HOUSE, TORRINGTON PLACE, W.C.1. LANgham 6633
MC273







STC

VALVES

FOR TELECOMMUNICATIONS, ELECTRONICS, AND INDUSTRIAL APPLICATIONS



RECTIFIERS • THYRATRONS • LOW AND MEDIUM POWER VALVES
TETRODES · TRIODES · BEAM PENTODES AND TETRODES

HIGH POWER VALVES
WATER COOLED · AIR COOLED · INDUSTRIAL TRIODES

MICROWAVE VALVES
TRAVELLING WAVE AMPLIFIERS · V M OSCILLATORS · KLYSTRONS

PULSE MODULATORS

COLD CATHODE VALVES
TRIGGER AND MULTI-CATHODE TUBES : VOLTAGE STABILISERS

VACUUM CAPACITORS • THERMAL DELAY SWITCHES
THERMOCOUPLES • BRIMAR TRUSTWORTHY RANGE
SUB-MINIATURE DIODES AND PENTODES

STC COMPONENTS GROUP

Standard Telephones and Cables Limited

Registered Office: Connaught House · Aldwych · London W.C.2

VALVE DIVISION: FOOTSCRAY · SIDCUP · KENT · FOOTSCRAY 3333

INDEX OF ADVERTISERS

Firm	page
ola Products Ltd.	xxxii
nec Ltd.	xxx
strong Whitworth Aircraft Ltd.	xv
omatic Telephone & Electric Co. Ltd.	xix
ing and Lee Ltd.	xi
ish Thomson-Houston Co. Ltd.	xvi
le Makers' Association	xxix
nodeon Crystals Ltd.	xxvi
a (A.R.L.) Ltd.	xx
vhurst and Partner Ltd.	
novan Electrical Co. Ltd.	
lish Electric Co. Ltd.	
lish Electric Valve Co. Ltd.	ii
Resistor Co. Ltd.	x
.G. Industrial Corp. Ltd.	xxxii
ranti Ltd.	i
K. Fox Ltd.	
neral Electric Company Ltd. (Telecommunications)	viii + ix
Johnson and Nephew Ltd.	xxx
lge Plugs Ltd.	
ndon Electric Wire Co. and Smiths Ltd.	xxxix
ude Lyons Ltd.	xxxiv
rconi Instruments Ltd.	xii
rconi Wireless Telegraph Ltd./A.T.E.	
rconi Wireless Telegraph Ltd.	vi + xxi
tropolitan Plastics Ltd.	
Metropolitan-Vickers Electrical Co. Ltd.	xviii
Mullard Ltd. (Magnetic)	xxxvii
Mullard Ltd. (Valves)	xxii
Newmarket Transistor Co. Ltd.	xvii
Newton Bros. (Derby) Ltd.	xxxiv
Oliver Pell Control Ltd.	vii
Philips Electrical Ltd.	xxxii
Plessey (Swindon) Ltd.	
Racal Engineering Ltd	xxiii
Rank Cintel Ltd.	xiii
Redifon Ltd.	iv, v + xxxv
Salford Electrical Instruments Ltd.	
Savage Transformers Ltd.	
Semiconductors Ltd.	xiv
Servomex Controls Ltd.	xxxi
Siemens Edison Swan Ltd. (Public Telephone Exchange)	xxvii
Siemens Edison Swan Ltd. (Public Telephone Apparatus)	xxxiii
Siemens Edison Swan Ltd. (with Redifon)	iv, v
Siemens Edison Swan	
Smiths Industrial Instruments Ltd.	iii
Standard Telephones and Cables Ltd.	xxiv, xxv, xxviii + xxxviii
Texas Instruments Ltd.	xl
Telephone Manufacturing Co. Ltd.	
Zenith Electric Co. Ltd	xxxiv

Wire for electrical progress

Lewmex or Lewkanex, Lewcosol or Lewcoglass—just a few of the 'LEWCOS' range of insulated wires designed to meet every need or application in the electrical industry, backed by intensive research and a first class technical service.



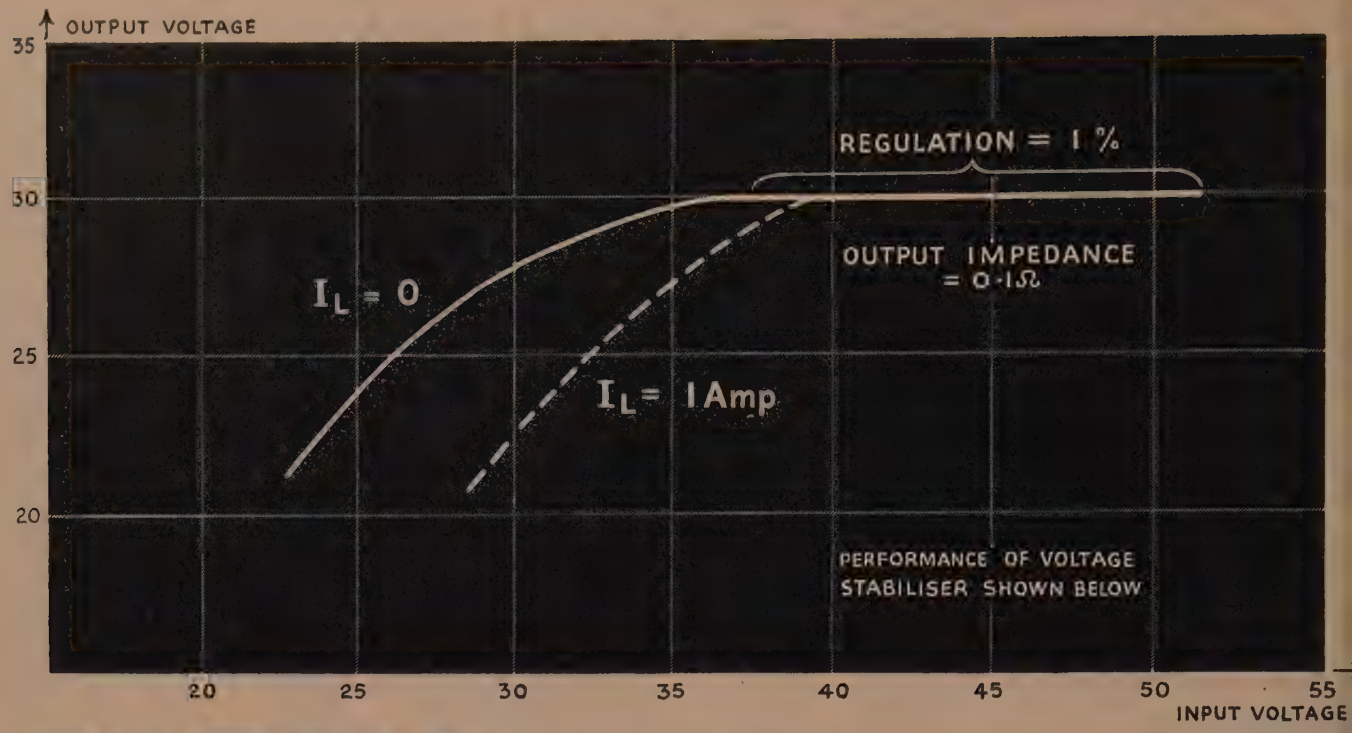
See us on Stand No. K10 at the Electrical Engineers Exhibition

LEWCOS

the largest manufacturers of insulated wires and strips in Europe

THE LONDON ELECTRIC WIRE COMPANY AND SMITHS LIMITED

LEYTON · LONDON · E10



NEW SILICON VOLTAGE REFERENCE DIODES

UP TO 10 VOLTS

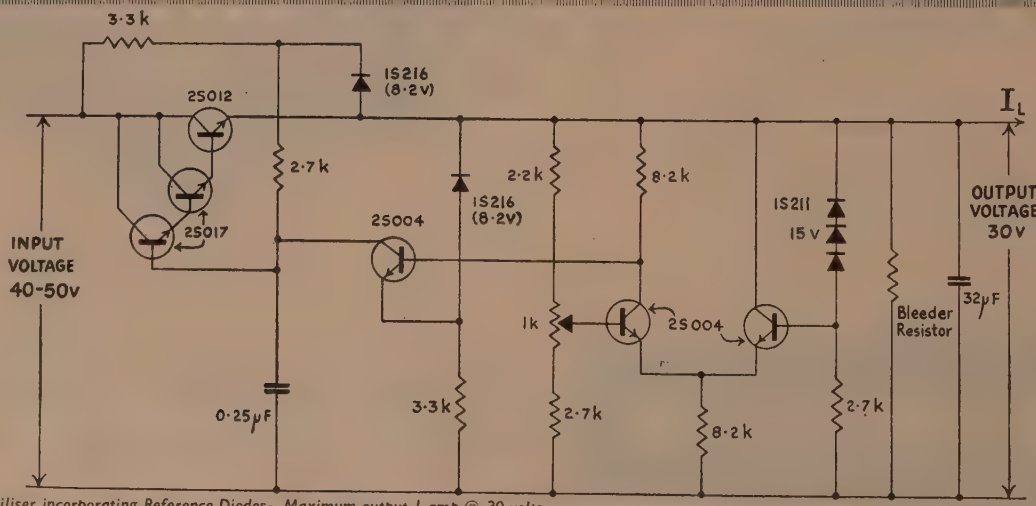
ACCURACY ±5%
(±10% Also Available)

OPERATING TEMPERATURES
− 65°C to +150°C

These glass-seal reference diodes cover Zener voltages from 3·6 to 10·0 volts in 12 steps, of approximately 10%. Performance data are shown for typical devices.

Type	Zener Voltage	Zener Current	Zener Impedance	Reverse Current		Power Dissipation		Typical Temperature Coefficient
	V_Z@I_Z volts	I_Z mA	Z_Z(max) @ I_Z Ohms	I_{Ib} @ −1V μA		P(max) mW		% / °C
				25°C	150°C	25°C	125°C	
1S207	3·6	20	22	0·15	20	400	80	−0·055
1S214	6·8	20	5	0·1	20	400	80	+0·035
1S218	10·0	20	12	0·1	20	400	80	+0·060





Voltage Stabiliser incorporating Reference Diodes—Maximum output 1 amp @ 30 volts.

Visit us on Stand No. S.3 at the A.S.E.E. Exhibition, Earls Court, 17th to 21st March.

TEXAS INSTRUMENTS LIMITED

TELEPHONE: BEDFORD 68051 **DALLAS ROAD BEDFORD** CABLES: TEXINLIM BEDFORD

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers.
An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF
HE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

L. 106. PART B. No. 26. MARCH 1959

395.5 : 621.395.74

The Institution of Electrical Engineers
Paper No. 2666 R
June 1958
©

RATING OF SPEECH LINKS AND PERFORMANCE OF TELEPHONE NETWORKS

J. SWAFFIELD, Ph.D., B.Sc.(Eng.), Member, and D. L. RICHARDS, B.Sc.(Eng.), Associate Member.

e paper was first received 4th November, 1957, and in revised form 12th February, 1958. It was published in June, 1958, and was read before the RADIO AND TELECOMMUNICATION SECTION *27th October, 1958.)*

SUMMARY

serious problem which occurs in the development of large or erwise costly speech communication systems (the term includes, but ot confined to, public telephone systems) is the achievement of a sistent overall design that is economical and yet never requires s to exert unreasonable amounts of mental or vocal effort in con- ation. The ideal thus simply expressed is notoriously difficult to eve and the problem is of long standing in the history of telephony. paper reconsiders this problem and sets out a coherent 'design' cedure by which studies of *rating* (providing figures of merit in ns that are meaningful for engineering design purposes) and *user uation* (study of merits of speech links for particular applications) be conducted to yield closely complementary information.

hese two aspects, 'rating' and 'user evaluation', are concerned ectively with individual speech links and with systems or networks ved as aggregates of such links. A further type of study, also ortant, deals with estimating the grade of performance which a plete network, designed to a given transmission plan, will provide; can be described in statistical terms such as the proportion of s of various types whose transmission should yield various levels satisfactoriness to users. This last type of study, conveniently wn as 'network transmission performance', is of special value in long-term planning and operation of large systems.

(1) INTRODUCTION

Many types of system for speech transmission over great or ll distances are in use; the paper is concerned with problems n of rating the effectiveness of their components for speech smission and of evaluating the usefulness of systems for ticular operational applications. The former leads to a sure, usually known as *transmission performance rating*, narily of interest within the design process, while the latter t in some way or other reflect user satisfaction in the opera- al context. Their combination into a coherent scheme that nits detailed engineering design and yields equipment that ts users' needs in the most economical way forms the central ure of the paper.

y virtue of its enormous size—over seven million telephones in the United Kingdom alone—and the variety of its equipment and users, the ordinary public network tends to dominate the scene. But the speculative element inherent in developments in other fields, such as defence telecommunications and some commercial applications, also demands confidence in transmission planning and hence in transmission performance ratings on which it is based and, indeed, presents some difficulties not present in more permanent networks of slow growth.

Systems, networks, etc., such as the above must be treated for transmission performance purposes as being composed of large numbers of individual 'speech links', i.e. sets of physical equipment and related factors (such as circuit noise) forming complete individual conversational paths. A speech link includes also ambient conditions at the ends, mainly acoustic noise, that affect transmission, but does not include the two particular subjects speaking over the link; use by average men is assumed. Thus the concept of 'transmission performance rating' (usually abbreviated to 'rating') is applied to individual links, whereas the value of an aggregate or 'population' of links of given ratings forming an operational system used for a definite purpose by a body of users is studied separately as a 'user evaluation' (see Section 6). This approach greatly assists practical studies; it does not, of course, avoid the difficulty that rating itself, being largely based on speaking tests of various kinds, involves measurements whose results depend on the human subjects taking part—their performances may differ widely from those of average men. This, however, is an experimental difficulty and must not confuse the meaning of the term 'transmission performance rating of a speech link'. For clarity the distinction will be retained of referring to people employed as material in rating tests as *subjects*, and to others, such as telephone subscribers or service personnel, forming the population who use equipment in the field (even though it may sometimes be as part of a user evaluation study) as *users*.

The interest of the paper is concentrated on systems whose purpose is the exchange of information by conversation carried on between two, or occasionally more, telephone stations. Broadcasting and public-address systems, in which essential elements of conversation are absent, are not considered. The type of use can vary from the fairly procedural type of speech

. Swaffield is at the Joint Speech Research Unit of the Post Office, Eastcote.
. Richards is in the Post Office Engineering Department.

1959: The Institution of Electrical Engineers

used on, say, air-ground control networks, where special training of users is essential, to the commercial and social conversations of the general public, where no training is possible.

Although the detailed work on which the paper is based is related to the speech transmission problems of the British Post Office and most of the examples are therefore of this type, an effort has been made to provide a framework for studies applicable over the wide field outlined above.

(2) TRANSMISSION PERFORMANCE

In spite of interest in transmission performance studies[1, 2, 3, 4, 18] dating from about 1905, methods have hitherto tended to remain narrowly related to specific types of equipment and use; results have consequently been of restricted application. However, by various expedients, many practical needs have been met. A telephone circuit can, for instance, be reliably rated relative to another fairly similar one. Thus a small telephone system can be designed and then tried in actual use. Slow, piecemeal extension of the system, including small variations of design, can be carried on, using, as planning information, ratings relative to equipment already in use. In this way, with its inherent element of user evaluation rather concealed, a large system can be built up provided its growth is so gradual that planning is only just running ahead of specific experience. The more quickly the system grows the more likely it is to acquire transmission failings which cannot subsequently be corrected at any reasonable cost.

Our present starting-point is the telephone conversation itself: it depends for its success, or lack of it, on the combined effects of three types of factors (conveniently designated as 'equipment', 'conditions' and 'users'):

X factors associated with the equipment, including the microphones and earphones. These may be regarded as controllable (by the designer).

Y factors associated with other physical conditions (mainly acoustical conditions at the terminal stations) affecting the transmission. These are regarded as uncontrollable.

Z factors associated with the users, including the actual use to which the speech link is being put. These are also uncontrollable.

The effects of the factors are strongly interdependent; for instance, the effect of a change of sidetone (an *X* factor) will differ greatly for different levels of room noise (a *Y* factor) and for different users (a *Z* factor). Furthermore, the range of the effects due to *Y* and *Z* factors commonly exceeds that due to the *X* (equipment) factors alone. General study must therefore be based on knowledge of what happens when all factors vary over a representative range, so that, among other things, the interaction of effects can be exhibited.

The problem can best be broken down by discussing separately two questions. The first concerns ratings of *X* and *Y* factors and is:

(*a*) On what scale of values, i.e. rating scale, is the transmission performance rating of individual speech links to be described? (It will be recalled that a link has been so defined that *Z* factors do not affect it.)

The most practical answer proves to be to compare performance with that of a Standard Speech Link containing a continuously adjustable parameter. The Standard Speech Link (often abbreviated to 'Standard') is made to a complete specification containing values for all *X* and *Y* factors; a single one of these is made adjustable and forms the variable parameter whose values provide the rating scale itself.

If these comparisons for rating measurements could always conducted with actual users and with full reproduction of ope tional conditions, perfectly valid results could, in principle, obtained. This course is impracticable for several reasons, o of which is the large number of detailed rating values neede The laboratory, or other technique of comparison which is to used instead, therefore becomes extremely important. This discussed elsewhere[46] and it will suffice here to note that suitable sample of subjects is necessary to ensure approximati to results for the average man embodied in the concept of rati a speech link, and also that the nature of the spoken exchang must have certain gross similarities to those taking place ov speech links. The function of speech links when embodied an operational system can then be left to question (*b*) below.

Rating techniques which have been described and extensive used include comparisons based on 'intelligibility', expressio of opinions of transmission quality, determinations of 'referen equivalent' (used by the C.C.I.T.T.), 'articulation' and 'immedia appreciation'. Practicability, precision and reliability are att butes influencing choice of method of comparison, but t critical feature is the extent to which realistic rating scales resu A good criterion for the merit of a rating process is the range links (of different types and of values of x) for which it remai true that every link rated as $x + \varepsilon$ is a better conversation connection than every link rated as x. Basically, it is limitati on range, in this sense, that calls for the methods (exploiti the interdependence of ratings, Standards and user evaluation described in the paper.

The second question concerns user evaluations based on study of *Z*-factor effects in operational use, and is:

(*b*) How is user satisfaction to be measured for speech lin whose performance has been rated in accordance with (*a*)?

Strictly, the evaluation is not of an individual speech link b of a working system that incorporates it; the distinction important because, commonly, the only manner in whic systems or networks can be described is statistically as a di tribution of (actual or potential) links of various ratings an types.

The answer to (*b*) prescribes the form of user evaluation Experience has shown that the opinions or behaviour of use during actual service, or operational trial, are the only reliab source of data for user evaluations; even then, satisfaction of t user is often more effectively indicated by his behaviour than his expressed opinion.

Attempts to simplify the overall procedure by partially com bining the studies of questions (*a*) and (*b*) have never prov very successful. If, as in the repetition-rate[5] type of study, measure too closely related to actual use is attempted in effort to make it also cover part of (*b*), then, apart from bei very slow to obtain, it tends to lack the precision needed f detailed design, while being still too formalized a measureme to be of much use for a user evaluation. Indeed, any studi which can be made only on a large working network can no more than indirectly useful for providing information guide further equipment design, since the results are averag over too many variable factors of all three kinds (equipmen conditions and users).

The danger in the procedure followed in the paper is th interactions between *Z* factors and any *X* or *Y* factor will concealed; the seriousness of this can be minimized by tw precautions: first by ensuring that the Standard Speech Link closely representative of practical field conditions (so far as and *Y* factors are concerned), and secondly by careful attenti to the technique by which rating comparisons between 'unknow and Standard speech links are made.

(3) METHODS OF RATING EQUIPMENT

(3.1) Transmission Standards and Standard Speech Links

Primarily the designer is interested in the factors (X factors) at he can control, such as overall loss, circuit noise, frequency aracteristic, etc. As already explained, the user effects (Z tors) may be segregated for separate study by rating speech ks in terms of a Standard Speech Link. Separation of X d Y factors from each other must be performed by subsequent alysis of the results of a programme of ratings.

Design can best proceed by aid of a physical circuit which has en applied in some practical situation—usually by being bodied in the design of a working system—and which has reby acquired a measure of user evaluation (if only by virtue absence of complaints, discussed in Section 6). Such a circuit termed a Transmission Standard and consists of equipment ch as microphones, earphones, amplifiers, line and exchange uipment; all its components are fixed and it represents only tures under the control of the designer (associated with X tors); it thus includes no specification of ambient noise.

In order to extend as far as possible the use of the simpler mparison methods mentioned in Section 3.4, a Transmission andard must be not too dissimilar to any circuits it is used to e. Where a convenient Standard is not available the gap ay often be bridged by use of a substandard which is then ed with great care relative to the Transmission Standard. ith experience in operational use a particular substandard will adually acquire the status of a Standard.

A typical composition of a Transmission Standard is that used the British Post Office; this is shown in Fig. 1 and further tails are given in Section 5.2.

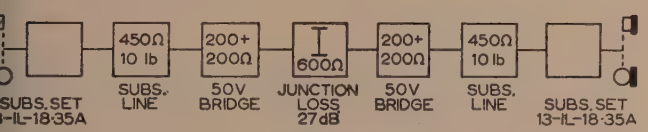


Fig. 1.—British Post Office Transmission Standard.

In order to use such a Transmission Standard for rating purses two features must be added to form a Standard Speech nk,* namely appropriate ambient conditions (usually simply om noise) must be specified, and a variable feature must be luded. The British Post Office, for reasons discussed in Section , has tended to use for room noise the rather high level of 60 dB continuous-spectrum noise shaped in frequency spectrum to nform to the average found in practice as described by Hoth.[10] e noise level is measured with the American Sound Level ter (70 dB weighting); the reading, so measured, corresponds proximately to the level of the noise above the threshold of libility.

Loss' or 'gain' is commonly included as the variable feature telephone rating problems, and where this is so it may be arded as modifying the fixed junction-loss of the Transmission ndard. Fig. 2 shows a typical form of rating measurement. e upper link (a) is a particular Standard Speech Link and the er is the unknown link to be rated. The terms 'near end' l 'far end' are assigned quite arbitrarily. The fixed loss x_1 ibels in the Standard Speech Link is part of the Transmission ndard (it corresponds to the 27 dB junction loss in Fig. 1). e variable feature is represented by N; the transmission con- on when $N = 0$ is described as 'datum transmission'. In case the fixed attenuator x_2 represents the attenuation of the

*Either of the terms 'Transmission Standard' and 'Standard Speech Link' may bbreviated to 'Standard' where there is no ambiguity or where distinction is nportant.

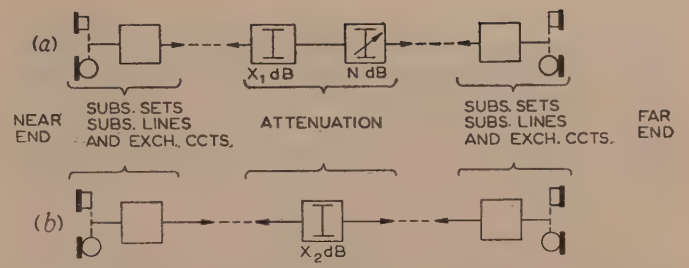


Fig. 2.—Rating of a speech link in terms of a link of comparable type used as a Standard.

(a) Standard Speech Link. It includes specification of ambient noise at both ends. Attenuation N is variable and may be + or −; $N = 0$ gives datum transmission.
(b) Link being rated. Ambient noise at both ends affects rating and must be stated.

trunk and junction network connecting exchanges in the unknown link.

The act of rating consists in balancing the two links to provide equal transmission performance; with the arrangement shown the rating of speech link (b) is then $-N$ decibels. Although only complete links can properly be rated, it is practicable to extend the meaning of ratings, first to ratings of individual components and second to 'sending' and 'receiving' ratings.

If only one component is different in the two links, a rating of $-N$ decibels may clearly be ascribed to that component alone and it may reasonably be described as 'N decibels worse than Standard'. Although there is no implication in the rating process that a link comprising components rated N_1, N_2, N_3, etc., will have the rating $N_1 + N_2 + N_3$, etc., this does prove to be empirically true for a wide range of combinations of components, and the use of ratings for transmission planning is obviously extremely difficult when this assumption cannot be made.

(3.2) Sending and Receiving Ratings

Separate 'sending' and 'receiving' ratings, though less easy to justify, are valuable, for example, because a change in sidetone affects sending and receiving by quite different mechanisms (the former because it influences talking level and the latter because it affects the level of reproduction of local ambient noise in the earphone). The near ends of the (a) and (b) links are arranged to be identical; the far end of the (b) circuit incorporates the change (most usually a microphone, earphone or sidetone circuit) to be rated. Measurement is conducted with speech first in the far-to-near direction and then in reverse, yielding settings N_s and N_r respectively. The sending and receiving ratings of the far end are now $-N_s$ and $-N_r$ respectively, and by the first extension of meaning these may be associated with the sidetone circuit, etc., directly if that is the only component that has been changed.

An interesting example of the value of an analysis based on sending and receiving ratings, as distinct from objectively measured characteristics of components, arises when increased sensitivities of microphone (say u decibels) and earphone (say v decibels) are considered. Owing to the effects of sidetone and noise, both of these affect both sending and receiving ratings. If w represents an increase in sidetone sensitivity (in this case better referred to as an improvement of $-w$ decibels) the following relationships have been found by extensive measurements of sending and receiving ratings. The relationships apply for the British P.O. Standard Speech Link whose details were given above (including the stated room noise):

$$-N_s = \text{sending rating} = u - \alpha(u + v + w) \quad . \quad . \quad (1)$$

$$-N_r = \text{receiving rating} = v - \beta(u + v + w) \quad . \quad . \quad (2)$$

where α and β are coefficients less than unity.

In these relationships, $(u + v + w)$ is of course the effective sensitivity of the air-to-air sidetone path. Eqn. (1) means that the sending rating does not improve as much as the microphone sensitivity because louder sidetone makes the subject lower his talking level; eqn. (2) means that, with the stated room noise, much of the improved earphone sensitivity is rendered useless by the corresponding increase in local noise transmitted via the sidetone path.

If equipment having these changed performances is brought into service, speech between two points will at any moment (whatever its direction) be affected by one of the sending and one of the receiving ratings. The rating of a complete link embodying the new components is therefore $-(N_s + N_r)$, and by adding eqns. (1) and (2) we have

$$N_s + N_r = (\alpha + \beta)w + (\alpha + \beta - 1)(u + v)$$

But $\alpha + \beta$ has been shown experimentally to approximate to unity, so that

$$N_s + N_r \simeq w$$

In other words, the final rating is practically independent of both the microphone and earphone sensitivities. This is of course true only over the range for which $\alpha + \beta = 1$; limiting factors to the range are ultimately the threshold of hearing masked by the talker's own speech or by leakage of noise past the earcap. This rather surprising result suggests concentrating on circuit improvements, such as better sidetone balance, that decrease sidetone sensitivity.

The separation of sending and receiving ratings for any given local end provides a practical solution to the troublesome rating problem of speech links whose transmission differs in the two directions, but it produces satisfactory results only when the level of performance is sufficiently high for non-conversational assessments to be valid.[46] Beyond this, however, the effects of both directions become increasingly entangled[29] and it becomes more appropriate to consider 'near-end' and 'far-end' ratings, i.e. ratings from the point of view of the subject near the item concerned and of the subject at the far end.

(3.3) 'Variable-Standard' and 'Margin' Methods of Rating

The discussion so far has assumed that the variable feature incorporated in the Standard is adjusted for balance and its setting at balance determines the rating of the unknown. For many purposes, however, an alternative method of rating in terms of 'margin' is preferable to the 'variable-Standard' method. Rating then consists in fixing the Standard to provide datum transmission ($N = 0$ in Fig. 2) and varying some form of degradation in the unknown until balance is achieved. The variable degradation may be circuit noise, line loss or some other feature of interest to the designer.

It can be shown that under the following conditions the variable-Standard and margin methods are equivalent and it becomes immaterial which is used:

(*a*) The degradation introduced into the unknown must be of the same kind as that forming the variable feature of the Standard.

(*b*) If Standard and unknown are equal at any given level of performance, a given additional amount of degradation in each must reduce the performance in such a way that they still remain equal at the lower level of performance.

These conditions also provide a satisfactory background to rating generally and should be met wherever possible. They provide another reason for care in erecting Standards and, if necessary, accepting a number of different Standards in orde to facilitate quick and reliable routine rating.

(3.4) Rating and Assessment

The term 'assessment method' is loosely used to cover tech niques (usually needing subjects) suitable for finding the poir of balance in Fig. 2. Irrespective of the actual subjective tech niques, e.g. articulation or conversational methods, there ar two ways of locating the balance: direct and indirect.

Direct balance is essentially a null method in which balance indicated but no value of performance or 'assessment score' attempted for either unknown or Standard. A simple and fairl satisfactory method is direct balance by conversational com parison. Two separate speech links are offered to a pair c subjects, one at the near and one at the far end; they convers over them alternately and subsequently indicate which lin seems better. The variable feature of the Standard Speec Link (or of the unknown if the margin method is used) is se at each of a number of settings fixed in advance and presente in turn in random order; decisions are recorded for each settin and the 'balance setting' can be estimated by an interpolatio method.[24] The method is particularly successful if the two link are sufficiently similar to cause no change in behaviour of th subjects, such as change of talker level. Conversationa behaviour takes 30 sec or more to become stable after a sub stantial change of quality, and it is quite impracticable to ge subjects to converse as long as this before recording ever decision. There are also severe restrictions on the absolut performance of the links for which this method is appropriate.

A very simple method, also based on direct balance, is th traditional 'loudness balancing'; this still provides the basis o which the transmission planning of international calls is dis cussed[36] (see Section 4.1).

An indirect-balance method overcomes some of the disad vantages of direct balance but is experimentally more cumbe some and needs altogether more vigilance if valid ratings are t result. For indirect balance the method of assessment mu provide a quantitative measure of performance, sometime known as a 'score', and balance is determined by equality o scores between unknown and Standard; several circuits can thu be rated in a single test in terms of a common Standard. Th main field of assessment studies is almost entirely concerne with methods providing a score and not with those useful merel in direct balancing. The range of levels of performance to b assessed, from those yielding almost perfect quality to those tha call for substantial lung power and concentration from th users, is a wide one over which conversational methods chang considerably and applicable techniques of assessment chang with them. For these and other reasons a variety of assessme methods exist[46] differing in many ways.

The relatively simple method known as articulation testin will sufficiently serve to illustrate the process of a scorin balance.* Each link is treated independently, though all sub jects, under a testing rota, probably speak and listen on a unknown links and on the Standard in turn. The articulatio score (proportion of test articulatory elements received correct in a formalized test)[25, 35] is plotted as a function of the variabl feature setting (usually attenuator setting in decibels) separate for all links including the Standard. This results in a fami of curves as shown in Fig. 3. When the level of performanc say 80% articulation, at which rating is required has bee decided, ratings can be determined from the horizontal intercep between the curves; the rating of Link 2 would be N_2 dB bett than the Standard as shown.

* The methodology of rating by a scoring method is formally the same as th developed in the field of biological assay[23] and indeed the history shows other remar able similarities.[9]

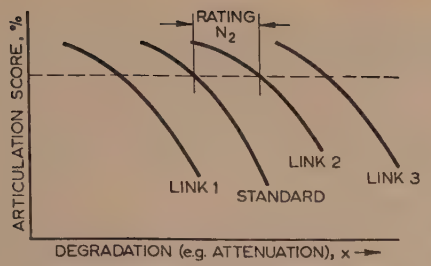


Fig. 3.—Rating by articulation—an example of indirect balancing.

The more primitive assessment methods have precision and eed but they lean, so to speak, heavily on the Standard for eir authority and must not be used to rate large departures om it. The most extreme combination of precision and nplicity at the expense of narrow range of application is the dely used 'objective' method based merely on gain and loss ferences at various frequencies.[14, 37]

RATING FOR TRANSMISSION PLANNING OF WORKING TELEPHONE SYSTEMS

(4.1) Function and Use of Transmission Standards and Reference Systems

Within a telephone administration and over a relatively short riod of time (say one or two decades), all telephone ratings n be conducted with the aid of a few Standards composed of actical items of equipment whose satisfactoriness has been monstrated in actual service. If the Transmission Standard one administration is to be related to that of another (a oblem that arises in planning international telephony) or if long-term transmission trends within an administration are be determined, some form of reference speech link, whose ecifiability and reproducibility are beyond question, is needed terms of which the various Transmission Standards can be ed.

It was exactly this problem that the C.C.I.F. set out to solve Paris in 1926. In 1927, at Como, the C.C.I.F. decided to up the European Fundamental Master Reference Telephone stem (S.F.E.R.T.) which was installed in Paris in the first f of 1928. This equipment, or 'system', was moved from ris to Geneva in 1948 and is still in use in the Laboratory of C.C.I.T.T. It was built of the highest-quality components n available, and contains, as variable degradation, an adjust- e attenuator between sending and receiving ends. A tech- ue of loudness balancing is specified to permit the rating of nmercial telephone circuits in terms of this system.[32, 36] tings obtained by means of this technique and equipment are med 'reference equivalents'. The equipment, although repro- cible at one stage in time, can hardly be so regarded any ger, and in any case its performance was not defined in terms t permit accurate reproduction now.

With modern equipment it is practicable to construct a high- ality reference system* whose performance is based on the ustical relationships which would exist in some ultimate ed) reference such as a short free-space air-path; some rticular value of the adjustable degradation can be arranged provide transmission identical with that of, say, a metre of

The system can be furnished also with means for adjusting sidetone sensitivity, noise level, etc. The basic features of h a system, for one-way transmission, have been described elsewhere;[20]* a complete reference speech link with both-way transmission needs two sets of this apparatus. Any reference system or Standard suitable for rating circuits over a wide range of performance must be both-way, although it may not always be necessary to use its both-way facilities when rating circuits of sufficiently high performance level.[46]

The high-quality form of reference system is unsuitable for use directly in day-to-day rating. It clearly does not meet the requirements for a Standard set out in Section 3.1; apart from transmission quality, the terminal conditions (e.g. a relatively distant-speaking microphone and headgear receiver on the one hand and a commercial handset on the other) usually make rating experimentally onerous and discourage the use of the reference system except for transmission comparisons between otherwise unrelated networks or systems and for the inter-rating of Standards and the study of rating itself.

(4.2) Transmission Performance Ratings

Transmission performance ratings are required for day-to-day design and planning and must therefore be capable of being easily understood and applied by staff responsible for installation of plant. This requirement is met in most telephone administrations (including the British) by framing installation instructions around a 'transmission limit', i.e. to the effect that the sum of the ratings of all plant involved in the worst connection must not be inferior to this limit. In practice, owing to the diversity of plant in use, some discretion must be exercised in deciding what plant constitutes the worst connection. Although not absolutely necessary, it is convenient if the Transmission Standard is so chosen that it represents the transmission limit, i.e. that its performance is the same as that of the installed plant for the worst connection.

Although the practice of setting up the Standard in two parts ('local ends' and 'junctions and trunks') has no fundamental significance, it is convenient in that it allows each part to be used separately at the ultimate installation stage. Rather different conditions apply to the local and the junction-and-trunk networks in practice; they spring ultimately from the fact that the latter is for the use of all while the lines of the former are exclusive to one, or at most two, users. The subdivision of the Standard also admits planning and development of the two corresponding classes of transmission equipment to proceed independently. British practice is similar in these respects to that of telephone administrations generally. The local part of the network can be altered only very slowly, and once a Transmission Standard has been set up and has become established in a working network, vast amounts of plant, both cable and equipment on users' premises, are planned and installed to corresponding limits.

As mentioned in Section 3.1, the British Post Office has, in forming a Standard Speech Link, frequently used its Transmission Standard in association with a room noise of a particular kind having a level of 60 dB. Choice of these rather adverse ambient conditions corresponds with the principle of making the Transmission Standard a representation of the installed plant involved in the worst connection; it follows that such poor *performance* as that represented by the Standard Speech Link set at zero (datum transmission) will practically never be reached in service. The effect of this is discussed in Section 5.

Complications arise owing to the need for clarity about monaural or binaural ning, one-way and two-way transmission, constitution of the ultimate reference m, etc. (see Reference 46).

* In 1949 a unidirectional high-quality system of this type was supplied by the British Post Office to the Laboratory of the C.C.I.T.T. at Geneva and is now known as A.R.A.E.N. (Reference Apparatus for the determination of Articulation Reference Equivalent).[33] When this equipment is set up with a 300–3 400 c/s band-pass filter and a certain level of circuit noise is injected into the receiving end, it forms a single-direction reference speech link termed S.R.A.E.N. (Reference System for the determination of Articulation Reference Equivalents).[34] Ratings, termed A.E.N. values (Articulation Reference Equivalents), are measured in terms of the variable loss between sending and receiving ends by a defined articulation technique carried out in the Laboratory of the C.C.I.T.T.

Although it is often appropriate to conduct the ratings at the level of performance defined by datum transmission, this is not always so, and substandards of different performance may be needed or different levels of room noise may be appropriate. In the examples which follow, a performance level of +14 dB has been used for rating local ends, whereas zero has been used in the circuit-noise example. Whatever the level of performance at which rating is desired it is essential that the assessment technique used in obtaining balance is itself appropriate to the level of performance. These considerations, which must be considered to apply to all ratings, whatever their purpose, are discussed elsewhere.[46]

The most important transmission-performance data required for planning a public network are concerned with the following:

(*a*) *For the local exchange network.*

(i) Maximum permissible lengths of subscriber's lines in various gauges of conductor (including mixtures of different gauges) used with all the available types of subscriber's sets and exchange feeding bridges.

(ii) Effects of introducing extension plans and P.B.X.'s.

(iii) Effect of idle pairs being teed on working lines.

(*b*) *For the junction-and-trunk network.*

(i) Effects of frequency-characteristic distortion, especially that arising from the use of unloaded and unequalized lines and from upper-frequency limitation due to heavy loading.

(ii) Effects of circuit noise arising either from the telephone plant itself or from sources outside, such as power systems.

(iii) Effects of transmission delay and echo.

The use to which data of type (*a*) are put is adequately described elsewhere;[16] the bulk of this detailed rating information is assembled by the simplified objective-measurement technique referred to in Section 3.4, but when different types of microphone or earphone are encountered, rating must depend on assessments using subjects. An example of rating a subscriber's set is given in Section 4.3.

Data of type (*b*) are not usually required in such detail as type (*a*), but the necessary rating measurements are much more difficult; an example dealing with circuit noise is given in Section 4.4.

(4.3) Rating a Subscriber's Set

When a new type of subscriber's set* is introduced into an established telephone network it is certainly not economically possible to replace every older type by the new set; the new will only gradually supplant the old over a period of years, and in the meantime it may even be very difficult to restrict the new (or the old) to certain classes of lines. For these reasons planning data for any new type of set must allow for the fact that connections will be set up with different types of equipment at the two ends. In practice obsolete types will remain in service possibly for decades, and more than two fundamentally different types must usually be catered for.

The most notable respect in which this difficulty arises is when a later-type set of increased sensitivity is to be introduced (for simplicity the earlier type will be assumed to be that embodied in the Standard). The increase in sensitivity is not usually distributed equally between the sending and receiving directions. As explained in Section 3.2, separate sending and receiving ratings can be extracted from ratings conducted on circuits with mixed sets. Together with the convenient but arbitrary rule that neither the sending nor the receiving rating may be worse tha its respective limit, this ensures that the sensitivity for neith direction of transmission of any connection in the network fa below that of the Transmission Standard. This concep though simple when sensitivity only is considered, becom increasingly unrealistic as more non-sensitivity factors, such noise, sidetone or distortion, are involved. One way round th difficulty is described below.

The separation of sending and receiving ratings clearly impli the use of a unidirectional method of assessment (see Section 3.2 It is explained elsewhere[46] that one-way methods are on applicable for rating speech links which are of sufficiently hig performance to give the listener no advantage from the retur path being available for requesting repetition, clarification, et The worst rating relative to datum transmission at which on way methods are applicable is about +14 dB, and the followin ratings have therefore been made using the listening-effo opinion-score technique described elsewhere[46] and the margi method of balancing against the Standard Speech Link set at rating of +14 dB. The arguments used below are therefo oversimplified in the sense that no consideration is given to th large numbers of calls that are not as much as 14 dB better tha Standard.

Any number of links tested in turn as unknown can be rate in this manner in terms of the Standard; each link consists of local end incorporating one type of subscriber's set connecte through a junction attenuator to another local end and possibl including a different set. In Fig. 4 the rating scale is shown (i) while the remaining scales show the corresponding junctio losses to achieve these ratings with various links. Scale (ii) giv the location of the junction-loss scale relative to the rating sca when two earlier-type sets (sets are associated with appropria

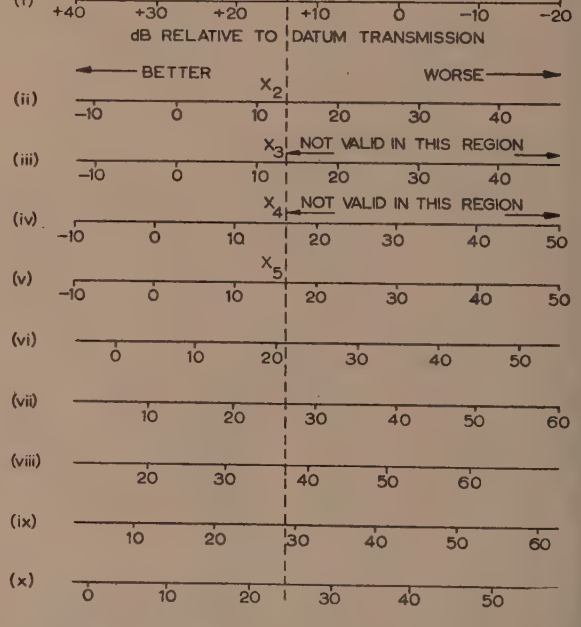


Fig. 4.—Transmission performance ratings for various speech links

(i) Rating scale.
(ii)–(x) Junction losses (yielding corresponding values on rating scale) of the follo ing links involving circuits given in Table 1:
(ii) Circuit (*a*) in 60 dB room noise.
(iii) Local end of circuit (*b*) sending to that of circuit (*a*) in 60 dB room noise
(iv) Local end of circuit (*a*) sending to that of circuit (*b*) in 60 dB room noise
(v) Circuit (*b*) in 60 dB room noise.
(vi) Circuit (*b*) in 50 dB room noise.
(vii) Circuit (*b*) in silence.
(viii) Circuit (*e*) in silence.
(ix) Circuit (*d*) in 50 dB room noise.
(x) Same as (ix) but with −60 dBm circuit noise at local ends.

* There appears to be no generally accepted term for the set of apparatus at the user's station. 'Subscriber's set' is used in connection with the public network (and is defined by the B.S.I.) but is obviously unsuitable for use with, say, air-ground communication or field use in the Army. The term 'telephone set' is used in the United States and may be generally acceptable.

Table 1

TYPICAL COMMERCIAL TELEPHONE CIRCUITS USED IN STANDARDS AND SUBSTANDARDS

Circuit	Near end		Far end	
	Telephone set	Subscriber's line (resistance and gauge)	Telephone set	Subscriber's line (resistance and gauge)
(*a*)	Earlier type (13–1L–18.35A)	Maximum (450 ohms, 10 lb)	As near end	
(*b*)	Later type (13–2P–27)	Maximum (660 ohms, 6½ lb)	As near end	
(*c*)	Later type (13–2P–27)	Maximum (660 ohms, 6½ lb)	Earlier type (13–1L–18.35A	Maximum (450 ohms, 10 lb)
(*d*)	Later type (13–2P–27)	Average (270 ohms, 6½ lb)	As near end	
(*e*)	Later type (13–2P–27)	Zero	As near end	

Note 1.—50-volt, 200+200-ohms feeding bridges apply in each case.
Note 2.—The local ends are connected together through a 600-ohm attenuator which forms the variable element.
Note 3.—Circuit (*a*) becomes the Transmission Standard when the junction attenuator is set at 27 dB.
Note 4.—Designations of the telephone sets correspond to explanation in Section 5.2.

cal lines) are connected together. Details of this circuit are ven as circuit (*a*) in Table 1, which includes a number of other cuits referred to below. Scale (iii) locates the loss scale for e later-type set, regarded initially as the sending end, connected the earlier-type set as the receiving end, i.e. circuit (*c*) in the ble transmitting near-end to far-end. Fortuitously, scales (ii) d (iii) have the same location relative to the rating scale. ale (iv) gives the loss for circuit (*c*) rated in the opposite rection, while scale (v) is that for later-type sets at both ends rcuit (*b*)]. This covers the four basic combinations of earlier d later sets.

Suppose that a rating 14 dB better than datum corresponds to s values x_2, x_3, x_4, x_5 on the four scales (ii), (iii), (iv), (v) pectively. A little thought will show that a sending rating the local end containing the later set can be defined as $+ x_5)/2 - (x_2 + x_4)/2 = 0$ dB, and a receiving rating as $+ x_5)/2 - (x_2 + x_3)/2 = +3$ dB. A corresponding inter-ion term $(x_2 + x_5)/2 - (x_3 + x_4)/2$ also exists, but is not ually significant.

The circuit shown as (*b*), containing only later-type sets, has is been fully rated and can now be regarded as a substandard d used for the rating, relative to it, of a large variety of ailed variations of circuits which include later-type subscribers' s; this detailed rating work can thus be done largely by simple thods (e.g. the purely objective method).

The effect of room noise other than the 60 dB specified in the ndard can be rated similarly. Scale (vi) in Fig. 4 shows the s corresponding to circuit (*b*) but with 50 dB room noise; le (vii) shows the same but with no room noise. The effect changing both subscriber's line and room noise is illustrated scale (viii), which shows circuit (*e*) with silent room conditions both ends.

(4.4) Ratings for Circuit Noise

Choice of a single, rather high, level of room noise has been tified for rating subscribers' sets on grounds of simulating her unfavourable conditions. However, if a rating of circuit se is attempted in the presence of this standard level of m noise, exceptionally small circuit-noise impairments, quite epresentative of average effects in practice, will result. The entive to make what are probably both practical and worth-ile reductions in circuit noise would thus on paper seem gligible. Realistic ratings of circuit noise should therefore be done with only modest amounts of room noise; 50 dB is a reasonable level.

As a simple example of a circuit-noise investigation it is best to consider conditions with average subscribers' lines rather than the lines of limiting length used in previous examples. Circuit (*d*) in Table 1 is composed of later-type sets and average lines; with a room-noise level of 50 dB and no circuit noise the location of the junction loss is shown in scale (ix) in Fig. 4; the injection of a specified type of continuous-spectrum circuit noise at a level of −60 dBm at the exchange displaces the loss scale to the position shown in scale (x). The movement in decibels of scale (x) relative to scale (ix) is the reduction of rating and is termed the 'circuit-noise impairment'.

The results of a whole programme of ratings of this type are summarized in Fig. 5, which shows, as a function of room noise and circuit noise, the junction losses that yield datum transmission performance. A circuit-noise level of −60 dBm in Fig. 5 corresponds roughly to the equipment noise likely to be contributed by long-distance trunk circuits.

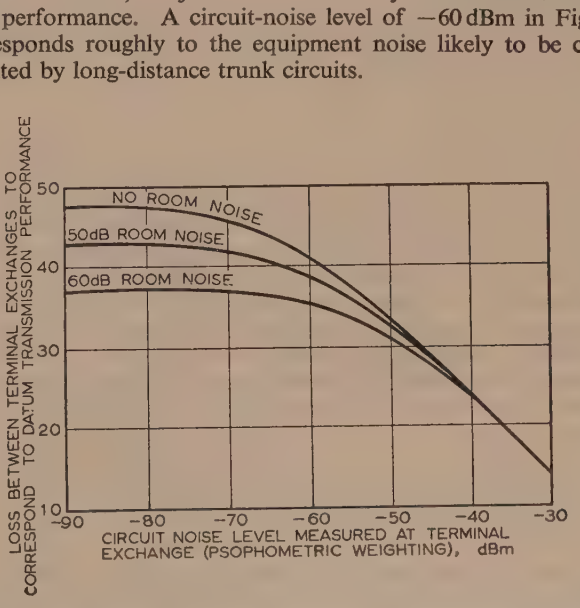


Fig. 5.—Loss between terminal exchanges to correspond to datum transmission performance as a function of room-noise and circuit-noise levels: circuit (*d*) in Table 1.

Room noise: continuous spectrum of typical shape.
Circuit noise: white, frequency band limited to 300–3 400 c/s.

(5) NETWORK TRANSMISSION PERFORMANCE

(5.1) The Problem of Planning a Network

How can the performance requirement of a whole system or network be stated in a form that can be usefully related to a transmission plan and then to installation instructions? The smallest units of usage in a telephone network are the individual conversations, and network transmission performance studies, as they are termed, are therefore concerned with estimates of the proportion of calls or conversations that would fail on transmission grounds. The causes of unsatisfactory performance are distributed between the X, Y and Z factors defined in Section 2, and so an obvious logical design procedure would be to use information from user evaluations (i.e. Z factors) to relate users' needs to ratings (i.e. Y and X factors), and then to relate ratings to details of equipment. This roughly corresponds to questions (*b*) and (*a*) of Section 2 in that order. This method cannot be carried out in practice because of the virtual impossibility of getting user evaluations of a kind suitable for design purposes, i.e. in precise form and explicitly related to ratings and design factors.

The practical order of procedure is as follows:

(*a*) Start with a rough guiding statement about the aim of the system. A quite general statement such as that 'nearly all connections should be satisfactory for conversation' may be sufficient to break into the process of design, construction and evaluation.

(*b*) A tentative Transmission Standard thought likely to achieve this aim can then be set up; with the distribution of the ratings of individual speech links relative to the Standard still to be discussed, precision is again unnecessary. This is discussed in Section 5.2 below.

(*c*) Detailed rating data relative to the Standard can then be acquired as discussed in Section 4.

(*d*) A transmission plan can next be formulated; this will give a distribution of the ratings of equipment, and hence the X factors which will apply on different calls; when associated with an estimated incidence of Y factors this yields a distribution of ratings relative to the transmission performance datum (Section 4.2).

(*e*) If such a distribution of ratings could be associated with data on the satisfaction of individual users associated with specific speech links (i.e. a distribution of Z-factor effects), it would be possible to determine an overall distribution of user satisfaction within the system (i.e. a measure of network transmission performance). It is generally impracticable to obtain Z-factor information in a suitable form; the most that can be expected is information on the satisfaction of a sample of users of a small aggregate of speech links. Nevertheless some simulation of Z-factor effects can be introduced by means of responses expressed by subjects in the course of controlled laboratory assessments such as those described elsewhere.[46]

(*f*) The final stage of the process is to compare the estimated network transmission performance with user-evaluation data. User evaluation tends to be a gradual process, often starting only when the system is in operation; though its analysis is not free from difficulty, it can yield information in a form that can be compared with that of the picture built up during stages (*a*)–(*e*) but now representing performance with users rather than subjects.

The penultimate stage (*e*) of the process provides two kinds of statement which taken together give a measure of network transmission performance. The first defines what is to be regarded as an unsatisfactory conversation; this will differ according to the intended usage of the network. Although a rigid definition is unnecessary it should be of sufficiently wide application to facilitate comparisons of network performance. One definitio is used in Section 5.3. The second kind of statement concern the distribution of unsatisfactory conversations in the network this is discussed in Section 5.4.

(5.2) Choice of Transmission Standard

In the design process it is the fusion of two quantities tha determines the service yielded by the network: the Transmissio Standard itself and the distribution of ratings in the networ relative to it. Given a free hand with the distribution, th Transmission Standard itself can, of course, largely be fixe arbitrarily, but it is convenient to assign to it some definit significance as is done by the practice in the British Post Offic of equating the Standard to the plant used in the worst class o calls, i.e. those between remote 'fringe' subscribers.* Th Standard, itself constructed of practical components, is later use as a basis for a body of detailed installation instructions to guid engineers in the field.

The initial erection of a Standard to represent the limitin transmission conditions is best done in the light of experience o similar networks, but in the absence of such experience it ca be done, exceptionally, by setting up the link in physical form and subjecting it to some kind of trial. If the equipment i fairly conventional it may be possible sufficiently to simulat operational use in the laboratory. In any case great care i needed if valid conclusions are to be drawn from trials.

The Transmission Standard used by the British Post Offic has evolved gradually over a long period. A Standard was firs set up about 1905, but a substantial network existed in som form long before then. The initial Standard, based purely o experience of what had in fact proved satisfactory, consisted o two subscribers' sets and lines connected by a length of artificia standard cable.[45]

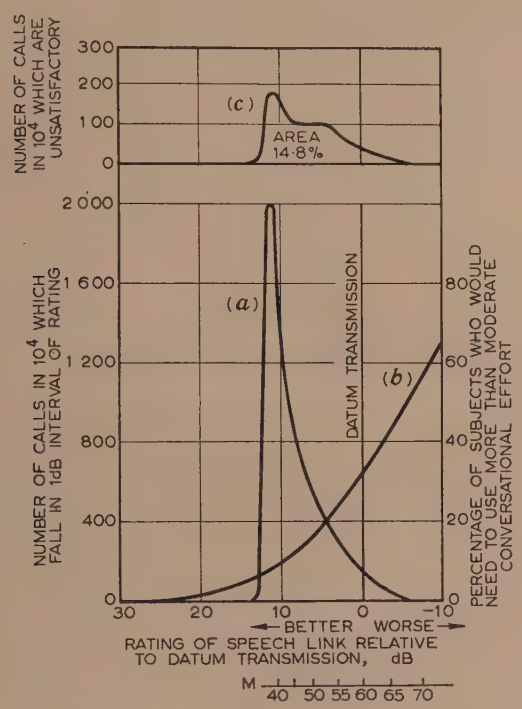
The Standard as revised in 1933 consisted of local end defined in terms of the then current style subscribers' sets an purely resistive artificial subscribers' lines, joined by 15 dB o line loss.[6] In 1946 the limiting local ends were again revise and more realistic methods of rating transmission performanc were also introduced.[14] This also made it necessary, whe calculating the transmission performance ratings of trunk an junction circuits, to include certain allowances for mismatc losses, switching-point losses and possibly frequency-ban limitation.[22] These additions augment the 15 dB loss fo the lines only, to the equivalent of about the 27 dB whic has been taken in the present paper. The Transmissio Standard in its present form is shown in Fig. 1. Th subscriber's set 13–1L–18.35A introduced about 1930 com prises a handset with a P.O. No. 13 transmitter capsule and P.O. No. 1L receiver, and a P.O. No. 18 induction coil associate with a P.O. No. 35A transformer, giving a small measure o sidetone suppression. With the addition of acoustic nois (60 dB in this case) at each end, the Standard Speech Link can b formed (Section 3.1) and used for rating; this automaticall locates the transmission datum (given by $N = 0$ in Fig. 2) fo use in more detailed planning.

(5.3) Distribution of Transmission Performance

Suppose for simplicity that the transmission plan of the tele phone network is effective in ensuring that calls of the wors type that has to be catered for (between remote fringe subscribers are in fact routed over plant equal in rating to the Transmissio Standard; this plant used with 60 dB of room noise at each en will yield datum transmission performance. More commonly lower levels of room noise will occur and 45 dB may be taken a

* A 'fringe' subscriber is defined as one who is connected to his local exchan by the worst permitted line and whose access to his Zone Centre is only through th maximum permitted number of switching points.

mean value; this is found to correspond to an improvement rating of 8 dB. Taking the distribution of levels of room oise present during different calls as represented by a Gaussian urve of mean 45 dB and standard deviation 10 dB, and using ie information given in Fig. 5, it is possible to calculate the istribution of ratings for the whole class of 'fringe-to-fringe' alls; this is shown by curve (*a*) in Fig. 6. The extreme skewness



ig. 6.—Illustration of the incidence of unsatisfactory calls as a function of rating, and calculation of a transmission-performance measure for a population of ratings applicable to 'fringe-to-fringe' calls.

(*a*) Distribution of ratings due to room-noise distribution (*Y* factors); circuit factors) corresponds to Transmission Standard.
(*b*) Subject effect (*Z* factors).
(*c*) Distribution of unsatisfactory calls.
Scale marked *M* gives room-noise level which yields rating shown when associated th the Transmission Standard.

esults from the fact that performance improves no further when ie room noise is reduced below about 40 dB whereas it deteriorates rapidly as the room noise rises above 50 dB.

So far, the equipment (*X* factors) and ambient conditions *Y* factors) only have been considered, but now the effects due the subjects themselves must be taken into account. To ake the example specific, assume that any subject who makes a all in which he needs to use no more than moderate conversaonal effort is enjoying 'satisfactory' performance and that when e needs to use more effort than this he is not enjoying satisctory performance. The onset of conversational effort and ie determination of what may be described as requiring no ore than moderate effort are described elsewhere.[46]* On this asis the percentage of subjects, as a function of rating, who il to enjoy satisfactory performance is reproduced in Fig. 6 curve (*b*).

Curve (*c*) is obtained by multiplying the number of calls given an ordinate of curve (*a*) by the probability, given by curve (*b*), at any of these calls would be unsatisfactory and gives the umber of calls per ten thousand which would be unsatisfactory a function of rating. The area under curve (*c*), 14·8%, is the tal proportion of unsatisfactory calls among all those whose

* Section 3.4 and Fig. 7 of the Reference quoted.

ratings are represented as a distribution by curve (*a*). The proportion of satisfactory calls is thus 85·2%, which may be taken as a measure of the grade of transmission service provided for 'fringe-to-fringe' calls considered as a class. It is extremely important to note that the greatest density of unsatisfactory calls occurs, not near the planning limit for plant (i.e. datum transmission), but at the high figure of about 12 dB better than this and near the rating for the more common calls. Thus the upgrading of plant rated near the lower limit would hardly affect the performance of the network as a whole.

Because curve (*b*) in Fig. 6 is of a form that is only asymptotic to zero, the grade of transmission service can never reach 100%, however much the Transmission Standard may be improved; in fact, if it could be improved by the wholly unrealistic amount of 20 dB, i.e. if curve (*b*) could be moved that amount to the right, this would still leave about 2% of these fringe-to-fringe calls unsatisfactory.

(5.4) Transmission Performance of a Class of Calls and of a Whole Network

The method illustrated below, which is completely flexible and can be applied to any types of call which form a convenient class for planning purposes,* consists in setting up a statistical model containing appropriate mean values and dispersion parameters for each variate and examining the model in the same way as the distribution of room-noise ratings was treated in Section 5.3; this leads to a measure of network performance for that class of calls in terms of the percentage which are unsatisfactory by the criterion used.

It would be possible to treat all calls that could be made in a network as a single population, but it is not very informative if the distribution of unsatisfactory calls in such a complex structure as a telephone network is thus reduced to a single numerical quantity. Separate statements are needed for various types of call, and in particular it is vital to know something of how unsatisfactory long-distance calls are distributed; if they represent a small proportion in the network as a whole but fall mainly to the lot of a few subscribers who are unfortunately situated the performance of the network cannot be considered as satisfactory. For simplicity, attention will be concentrated here on long-distance traffic; if this service is reasonable, that provided on local and short-distance calls will usually also be acceptable. However, a network in which very few long-distance calls were in fact made might justify a more stringent criterion of satisfactoriness for local calls.

The following example is in terms of long-distance calls on the British Post Office network, though it does not refer to the exact situation at any particular time. Defining long-distance calls as those which involve at least two Zone Centres, it is possible to determine the statistical distribution of their ratings. This will, for argument, be represented as a Gaussian curve of mean rating +28 dB relative to the transmission performance datum, and standard deviation 5 dB including the effect of room noise. Performing the same computations as described in Section 5.3 (and there illustrated by Fig. 6) yields the percentage of all long-distance calls which are satisfactory by the criterion used in Section 5.3 as over 99%.

A fringe subscriber will suffer a lower grade of transmission service than average, and, making reasonable assumptions regarding the distribution of ratings for long-distance calls originated by a fringe subscriber, about 97% of such calls will

* Types of call in this sense are naturally associated with the switching arrangements of the network. In the United Kingdom network the main switching points, apart from local exchanges, are known as 'Group Centres' and 'Zone Centres'. The United Kingdom is divided into about 15 zones most of which contain perhaps 10–20 groups (though some have more and some less). Thus the most general form of long-distance call will involve a Zone-to-Zone link and its path will be of the form: (subscriber) → (exchange) → (Group Centre) → (Zone Centre → (Zone Centre) → (Group Centre) → (exchange) → (subscriber).

Table 2

TYPICAL TRANSMISSION PERFORMANCE OF A TELEPHONE NETWORK

Type of calls	Approx. no. of calls per week in the class (B.P.O. network April–June, 1954)	Proportions of calls by various criteria of 'satisfactoriness'		
		No listening effort needed (roughly 'excellent')	No conversational effort needed (roughly 'good' or better)	No more than moderate conversational effort needed (roughly 'fair' or better)
		%	%	%
(1) All calls involving at least two Zone Centres	1 600 000	89	95	99·8
(2) All calls involving at least two Zone Centres and one fringe exchange*	26 600	64	86	99
(3) As (2) but subscriber on fringe exch. has near-limiting exch. line	2 000	50	80	97
(4) Calls between two remote fringe exchanges*	700	47	79	97
(5) As (4) but both subscribers on near-limiting exchange lines†	4	About 10	51	85

* A 'fringe' exchange is one whose access to its Zone Centre is through the greatest permitted number of links.
† 'Near-limiting exchange line' here means greater than 0·85 of the limiting line length; such lines usually comprise about 7½% of the total.

be satisfactory by the same criterion. Calls between remote fringe subscribers will naturally be worse still, and about 85% would be found satisfactory.

By employing other criteria[46] of satisfactoriness a complete picture of the performance provided by the network as a whole can be constructed. Table 2 shows some more extensive results for five types of call.

The above description of the performance of a network must be viewed in its context as part of the design and planning process; in itself it provides little absolute information about whether the users are likely to be satisfied with their service. This can only be obtained from operational studies of some kind such as are indicated in Section 6. Changes in the detailed picture provided by a set of figures of the kind given in Table 2 provide an extremely sensitive indication of likely changes in network performance due to any contemplated change in network planning policies; in a large network these are often very difficult to foresee. Studies of the kind illustrated enable the present performance of a class of calls or of a whole network to be estimated, and provide evidence for confident statements such as that the network would not deteriorate as a result of some proposed change of transmission planning, or that such-and-such an economy could, without lowering the overall aim, gradually be achieved if it were done by degrading the performance in one particular respect in which it was proving unnecessarily good.

By elaborating the statistical model more complicated problems may be studied. The method has been applied, for instance, to a problem where, in addition to factors already considered, the circuit noise could vary widely from one call to another owing to slow fading of a radio relay link.[44]

(6) USER EVALUATION OF SPEECH LINKS AND TELEPHONE NETWORKS

In conclusion, the relatively small but essential work of 'user evaluation' of equipment must be outlined; in effect, as has been seen, it is user evaluation of the chosen Transmission Standard taken together with the method of applying it in installation. General acquaintance with the system in service and the extent of praise or complaint it generates can sometimes yield sufficient indication that the network is satisfactory, though it is unlikely to guide it towards the most economical form.

Where speech equipment has to meet some fairly narrowly defined operational requirement and users have specific duties and appropriate training (e.g. short-range air-ground communication for aircraft control) the equipment may be subjected to straightforward user trials; two points should be noted, as such trials often have to serve a much wider purpose than mere user evaluation of the transmission aspect of the equipment:

(*a*) Trials must be so conducted that conclusions can be drawn specifically about the Transmission Standard around which the design has been worked out.

(*b*) Achievement of (*a*) is unlikely unless the equipment is made to perform its full function in operational service and is not merely used for speaking trials.

Provided these points are borne in mind the trials present problems similar to other types of user trials.

A different situation is presented by general-purpose telephone networks, especially where the users are members of the general public. Three methods of approach are suggested. All are practicable and their merits are complementary; taken as a whole they could be made to form the basis of a sound continuing evaluation.

The first method is simply the maintenance of a continuous watch on the level of public satisfaction; it is largely an administrative matter and provides only a general guidance whose main criterion will be change in the nature or volume of public complaint. Vague as it necessarily is, this form of information is the most direct source of confidence in the Transmission Standard used and in the method of applying it.

The second method relies on voluntary co-operation from subscribers, a very small proportion of whom would be asked to report (orally or possibly by answering questionaries) on transmission conditions on their own telephones in *normal use*. The inquiry must be very carefully designed if it is to yield information on the satisfactoriness of calls for the purposes for which they were made, but the difficulties can be largely overcome and this type of method has been highly developed in comparable fields.[12]

The third method takes the form of transmission and noise measurements by testing officers between pairs of telephones selected on some sampling basis. Essential features of the method are that (*a*) measurements would be conducted from actual subscribers' stations and not from exchanges or other line terminals, and (*b*) the stations would be selected on a country-

de sampling basis. The results would not directly measure er satisfaction and could therefore be used only in conjunction th those of the second method. Nevertheless, with experience d with suitable precautions, good estimates of proportions of isfactory calls should be possible.

The basis on which the adequacy of the performance of a blic network may be assessed may now be summarized. The stem as a whole may be judged as broadly satisfactory if there a low level of public complaint over a long period. This in rn implies roughly that the Transmission Standard and the sociated distribution of ratings relative to it are satisfactory. mparisons can then be made, in terms of proportions of satisfactory calls, between the *planned* performance distribun of the type shown in Table 2 and the corresponding achieveent obtained from the sampling surveys just outlined. Even the network is entirely free of fault conditions no close solute agreement in these comparisons is to be expected, for o reasons: (*a*) while careful choice of assessment methods will inimize differences, the precise extent to which user satisfaction service corresponds to favourable opinions of subjects under ntrolled conditions is necessarily unknown; and (*b*) whereas factors are automatically taken into account in a user evaluan, it may not be possible to locate and include all factors network performance calculations. However, any residual ctors causing discrepancies are likely to be fairly constant in ch case, and careful initial study and firm adherence to produre should, in the course of time, enable the ratio (or changes the ratio) of achieved to planned proportions of, say, 'unsatisctory calls per million' to yield a useful indication of the state maintenance of the network as a whole. Any mechanism of e order of size of a large network—hundreds of millions of mponents—must function with a certain proportion of comnents unserviceable or not to specification; this degrades the erall performance of the network and must be considered ng with, but not in preference to, other degrading factors. owever, a means of measuring maintenance level as it affects nsmission has not hitherto been available.

Study of the distribution of unsatisfactory calls per million er various types of connection can also show what must be ne to reduce the occurrence of lines bearing an unreasonable are of transmission troubles, and in other ways assist analysis the performance of the network.

Finally, user evaluation of the form discussed here would also able the actual performance of the network to be recorded r comparison with that at earlier times, and would show ether improvement could more profitably be sought in plang or in maintenance. It is only in the light of these long-term d rather complicated factors that a full evaluation of a Transssion Standard as embodied in a telephone network can adually be established.

(7) CONCLUSIONS

So far as is known, no previous attempt has been made to ify and reduce to systematic form the processes of rating dividual speech links and appraising the transmission aspects whole systems. This task, which the paper attempts, proves sively difficult even where the plant involved is on a sufficient le to justify considerable investigation. The paper does not present any single investigation but is the outcome of work rsued at intervals during a number of years.

In the end it has turned out that an inductive rather than a ductive approach is the more practical. A particular speech k, conveniently known as the Transmission Standard, is used represent a tentative standard for transmission planning. This ndard is then used to connect together 'ratings' (as used by e designer) and 'evaluations' (as made by the body of users); in effect the Standard is the means of accumulating relevant experience as it is acquired, and thence of strengthening confidence for the planning of a more extensive network.

A secondary purpose of the paper is to offer some common ground between the various fields concerned, including that of public telephone systems, and so to further the development of sound design methods applicable to large speech-transmission systems.

(8) ACKNOWLEDGMENTS

The authors are indebted to many present and former colleagues at the Post Office Research Station for their help, especially in the course of prolonged experimental work. Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of information contained in the paper.

(9) BIBLIOGRAPHY

(1) CAMPBELL, G. A.: 'Telephonic Intelligibility', *Philosophical Magazine*, 1910, **19**, p. 152.

(2) CRANDALL, I. B.: 'Composition of Speech', *Physical Review*, 1917, **10**, p. 74.

(3) FLETCHER, H.: 'The Nature of Speech and its Interpretation', *Journal of the Franklin Institute*, 1922, **193**, p. 729.

(4) FLETCHER, H., and STEINBERG, J. C.: 'Articulation Testing Methods', *Bell System Technical Journal*, 1929, **8**, p. 806.

(5) MARTIN, W. H.: 'Rating the Transmission Performance of Telephone Circuits', *ibid.*, 1931, **10**, p. 116.

(6) STRATTON, J., and LUXTON, W. G.: 'Modern Developments in Telephone Transmission over Lines', *Institution of Post Office Electrical Engineers, Printed Papers*, No. 153, 1933.

(7) GRINSTED, W. H.: 'The Statistical Assessment of Standards of Telephone Transmission', *Engineering Supplement to Siemens Magazine*, January, 1937, No. 140, p. 1.

(8) HUGHES, J. R.: 'Progress in the Study of Effective Transmission by "Immediate Appreciation" Tests and Statistical Assessment', *ibid.*, November, 1938, No. 162, p. 1.

(9) FIELLER, E. C.: 'Biological Standardization of Insulin', *Journal of the Royal Statistical Society*, Supplement, 1940–41, **VII**, No. 1, p. 1.

(10) HOTH, D. F.: 'Room Noise Spectra at Subscribers' Telephone Locations', *Journal of the Acoustical Society of America*, 1941, **12**, p. 499.

(11) HUGHES, J. R.: 'The Comprehensive Assessment of Telephone Communication Efficiency', *Journal I.E.E.*, 1942, **89**, p. 195.

(12) SILVEY, R. J. E.: 'Methods of Listener Research employed by the British Broadcasting Corporation', *Journal of the Royal Statistical Society*, 1944, **107**, p. 190.

(13) 'American Standard Sound-Level Meters for Measurement of Noise and other Sounds', American Standards Association, Specification Z24.3, 1944.

(14) BARNES, E. J., WOOD, A. E., and RICHARDS, D. L.: 'Standards of Transmission and Local Line Limits', *Post Office Electrical Engineers' Journal*, 1947, **39**, Part 4, p. 151, and **40**, Part 1, p. 8.

(15) POCOCK, L. C.: 'A Survey of the Telephone Transmission-Rating Problem', *Journal I.E.E.*, 1948, **95**, Part III, p. 253.

(16) FUDGE, G. A. E.: 'Introduction and Application of Transmission Performance Ratings to Subscribers' Networks', *Institution of Post Office Electrical Engineers, Printed Papers*, No. 198, 1949.

(17) SHANNON, C. E.: 'A Mathematical Theory of Communication', *Bell System Technical Journal*, 1948, **27**, pp. 379 and 623.

(18) EGAN, J. P.: 'Articulation Testing Methods', *Laryngoscope*, 1948, **58**, p. 955.

(19) RADLEY, W. G.: 'Speech Communication under Conditions of Deafness or Loud Noise', *Journal I.E.E.*, 1948, **95**, Part I, p. 201.

(20) SWAFFIELD, J., and DE WARDT, R. H.: 'A Reference Telephone System for Articulation Tests', *Post Office Electrical Engineers' Journal*, 1950, **43**, p. 1.

(21) RICHARDS, D. L.: 'An Outline of the Design and Analysis of Experiments', *ibid.*, 1951, **44**, p. 55.

(22) Contribution of the British Telephone Administration to the study of Question 9 of the 4th Study Group. (Transmission Performance Rating.) C.C.I.F. Yellow Book 1951, **1** ter, Annex 5 to Question 9, p. 396.

(23) FINNEY, D. J.: 'Statistical Method in Biological Assay' (C. Griffin and Co. Ltd., London, 1952).

(24) RICHARDS, D. L.: 'Design and Analysis of Subjective Acoustical Experiments which involve a Quantal Response', *Acustica*, 1952, **2**, p. 83.

(25) DE WARDT, R. H.: 'The Conduct of Articulation Measurements', *Post Office Electrical Engineers' Journal*, 1952, **44**, Part 4, p. 159.

(26) RICHARDS, D. L., SHEARME, J. N., and BLUNDEN, R. M.: 'A Method of Generating Random Shapes for use as Material in the Message Rate Method of Assessing Telephone Circuits', Post Office Research Report No. 12970, 1952.*

(27) RICHARDS, D. L., and BERRY, J.: 'Tables for the Generation of Logatoms based on Conversational Speech', Post Office Research Report No. 13596, 1952.*

(28) FLETCHER, H.: 'Speech and Hearing in Communication' (Van Nostrand, 1953).

(29) RICHARDS, D. L.: 'Some Aspects of the Behaviour of Telephone Users as affected by the Physical Properties of the Circuit' ('Communication Theory', Willis Jackson, Ed.), p. 442 (Butterworths Scientific Publications, 1953).

(30) RICHARDS, D. L., and SWAFFIELD, J.: 'Measurement of the Communication Rate of a Speech Link—a Preliminary Appraisal of the Problem' ('Communication Theory', Willis Jackson, Ed.), p. 450 (Butterworths Scientific Publications, 1953).

(31) GUILDFORD, J. P.: 'Psychometric Methods' (McGraw-Hill, 1954).

(32) 'Reference Systems for the Determination of Referenc Equivalents', C.C.I.F. Green Book, 1954, **4** (3.1.1) an (3.1.2), p. 27.

(33) 'Reference Apparatus for the Determination of A.E.N (A.R.A.E.N.)', *ibid.*, **4**, Section B, p. 56.

(34) 'Description and Adjustment of the Articulation Refe ence System (S.R.A.E.N.)', *ibid.*, **4** (3.1.3), Section C p. 65.

(35) 'Measurement of A.E.N. of a Commercial Telephon System (sending and receiving) by Comparison with th S.R.A.E.N.', *ibid.*, **4** (3.1.4), p. 68.

(36) 'Measurement of Reference Equivalents and Relativ Equivalents', *ibid.*, **4** (3.4.2), p. 135.

(37) 'Summary of a Method used by the British Telephon Administration to Determine Relative Transmissio Performance Ratings of Local Sending and Receivin Systems using the same Types of Microphone an Receiver', *ibid.*, **4**, Annex 2.

(38) 'Immediate Appreciation Testing Method', *ibid.*, Annex 23.

(39) 'Method used by the British Telephone Administration fo the Subjective Assessment of Telephone Transmissio Quality', *ibid.*, **4**, Annex 24.

(40) WOODWORTH, R. S., and SCHLOSBERG, H.: 'Experimenta Psychology' (Methuen, 1955).

(41) QUASTLER, H.: 'Studies of Human Channel Capacity ('Information Theory', Colin Cherry, Ed.), p. 36 (Butterworths Scientific Publications, 1956).

(42) RICHARDS, D. L., and ARCHBOLD, R. B.: 'A Developmen of the Collard Principle of Articulation Calculation' *Proceedings I.E.E.*, Paper No. 2143 R, September, 195 (**103** B, p. 679).

(43) Contribution of the Telephone Administration of th U.S.S.R. to Question 7 of the 4th Study Group, C.C.I.F 1955–6, 4th S.G.—Documents No. 26 and 36.

(44) 'A Statistical Study of Circuit Noise with a Specified Tim Distribution', C.C.I.F., 1955–56, 4th S.G.—Documen No. 39, p. 34. (Reproduction of Post Office Researc Report, No. 14069, 1956.)*

(45) 'Transmission and the Trunk Network', *Post Office Elec trical Engineers' Journal*, 1956, **49**, Part 3, p. 21 (especially pp. 223–225).

(46) RICHARDS, D. L., and SWAFFIELD, J.: 'Assessment o Speech Communication Links', (see next page).

* Restricted report, not generally available.

* Restricted report, not generally available.

[The discussion on the above paper will be found on page 90.]

1.391 : 621.395.5 : 621.395.813

The Institution of Electrical Engineers
Paper No. 2605 R
Apr. 1958
©

ASSESSMENT OF SPEECH COMMUNICATION LINKS

y D. L. RICHARDS, B.Sc.(Eng.), Associate Member, and J. SWAFFIELD, Ph.D., B.Sc.(Eng.), Member.

he paper was first received 21*st November*, 1957, *and in revised form* 10*th February*, 1958. *It was published in April*, 1958, *and was read before the* RADIO AND TELECOMMUNICATION SECTION 27*th October*, 1958.)

SUMMARY

Testing methods, such as 'articulation', 'immediate appreciation', c., at present available for providing a direct measure of the per-rmance or effectiveness of a link for speech communication are not dividually or even collectively applicable over any wide ranges of vels of performance or types of link. Many practical links fall tside the range over which these methods are reliable, and there is ed for a general method that will both provide a safeguard against realistic results and serve, if possible, to extend the usefulness of isting techniques.

The paper examines the mechanism of conversation over a speech nk and deduces a sequence of criteria each applicable over an appro-iate range of performance. A general assessment method follows om this, and the corresponding experimental techniques (which epend on the use of ordinary untrained subjects) and some applica-ons are discussed. Some existing assessment methods can be fitted to the framework provided by the general method and can then be ed with increased confidence.

Assessments as described in the paper are of little value for indicating e merit a link has for the user. Their most valuable use is to facilitate s part of the process known as rating) realistic comparisons of links ith Standard Speech Links whose evaluation can be studied in tual use.

(1) OBJECTS OF ASSESSMENT

An assessment of the performance of a speech link is the sult of some direct measure of its performance or effectiveness r speech communication. The measure, variously referred to an 'assessment score', an 'assessment' or a 'score', may be some such quantity as 'fraction of material intelligible' or, ore simply, it may be some statistic describing the opinion of bjects who have used the link for conversation. When used ithout qualification, as here, assessment usually implies results btained directly from tests using subjects. In the case of ticulation scores[26] it is possible to make calculations instead conducting actual tests, and this is sometimes useful; calcula-on is, however, only a means of using indirectly results from bjects, since the data for calculation are entirely empirical and fundamental method of calculation is possible.

A wide range of subjective techniques exists and the conduct the various types of tests has become highly developed. They ry from extremely formalized tests, such as articulation, of latively high precision, to more cumbersome and less precise sts based on the exchange of conversational material between rge numbers of untrained subjects. Laboratory tests of this nd are part of the apparatus of 'rating' one speech link lative to another (usually an 'unknown' link relative to a andard Speech Link*) and this is one of their principal uses. he present paper is, however, concerned only with assessments emselves; the purpose and application of rating methods are scussed in another paper by the same authors.[28] Essentially is individual complete links that are assessed rather than her components of links or links as elements in systems or networks; the use of ratings as distinct from mere assessments is essential for the study of these latter aspects.

Whatever they might ideally do, methods of assessment that can be conducted in practice are thus of very limited direct value either to the user of a network, since they do not tell him how useful a link is for his purpose, or to the system designer, because his interest is necessarily in ratings. Assessment methods are, however, indirectly vital for both these purposes. The uses of assessment methods are well defined and worth enumerating; they are set out below roughly in order of increasing difficulty in conducting them effectively:

(*a*) For assessing, by direct balancing, equality of transmission performance between a Standard Speech Link (usually abbreviated to Standard) and 'unknown' in rating.

(*b*) The same as (*a*) but by indirect balancing.

(*c*) For providing a performance specification of a Standard in much simpler and more specific terms than does a user evaluation. As much information as possible is wanted about Standards that are to be widely used.

(*d*) As a rough yardstick for research projects concerned with new types of speech transmission when it is not practicable to compare with a suitable Standard or when special features of research interest are under detailed study. The performance description (e.g. proportion of spoken material immediately appreciated) can be pressed into service as an absolute measure.

(*e*) As a useful guide when speech equipment whose development is not yet complete has to be considered for some new operational use for which no applicable physical Standard exists and user trials are not yet possible. When these circumstances occur it is, however, usually better either to attempt some condensed and specialized user trial or to use whatever experience is available towards setting up a tentative Standard. There is little merit in arguments that deduce the practical value of a system in operation directly from, say, sentence intelligibility scores under experimental conditions.

The essentials of an assessment scheme are the same whatever use is to be made of the results. Methods suitable only for use in detecting equivalence of performance in the elementary rating process known as direct balancing [(*a*) in the above list] are, however, mentioned in Section 3.4 of the paper[28] already cited and will not be further discussed here. Of the remaining methods, with which the present paper deals, the extent to which they are of any value as absolute methods varies considerably; this, however, has only a secondary influence on their application to rating by indirect balancing [(*b*) in the above list], which is the most important application.

Except for those simplified forms only suitable for detecting equivalence, an assessment deals with a single speech link at a time. For the purpose of description there is thus no loss of generality, as between the applications (*b*) to (*e*) above, in considering the performance of a single link. (When the process is applied in turn to both Standard and unknown in rating, very short tests on each are minutely interleaved for experimental reasons though scored separately.) The foundation of any

* Various terms, including 'speech link' and 'Standard Speech Link' are explained Section 8. The actual uses of Standard Speech Links are discussed elsewhere;[28] efly they comprise working-type telephone apparatus, representative of equipment ervice, associated with specified ambient conditions at both ends.

Mr. Richards is in the Post Office Engineering Department.
Dr. Swaffield is at the Joint Speech Research Unit of the Post Office, Eastcote.

scheme of assessments is a set of performance criteria related to the way speech links of widely varying merit are used in conversations. The mechanism of a link as used for conversation is considered below before various practical assessment methods are discussed.

(2) THE MECHANISM OF A SPEECH LINK AND CRITERIA FOR ITS PERFORMANCE

The nature of a telephone conversation is considered below in some detail to explain the intractable nature of the problem of measuring its performance; in Section 3 the results are applied in devising a method of assessment, intimately related to the varying 'mechanism' of a conversation, to suit speech links of different performance levels. Plausible explanations of poor performance of a link and reasons for unrealistic, unrepeatable or, all too frequently, anomalous results of assessments yielded by particular methods are relatively easy to devise; it is far less easy to predict these things and show the way to avoid them.

(2.1) Communication over Channels of Different Levels of Performance

The temptation to measure some quantity that can be identified with 'fractions of information correctly received' is very strong. It is, however, rather analogous to rating a transport service by the criterion 'proportion of goods received undamaged'; a high assessment is a necessary but hardly a sufficient achievement and could at best serve to discriminate between a number of alternative services all so bad that the fraction of goods damaged in transit was a dominant aspect of the service. Even if the service is destructive in this way a better measure of its value might be the amount of extra effort, e.g. packing, needed to overcome the trouble. If, however, the service *can* convey more or less undamaged the goods to be sent, the user's interest is likely to be transferred to quite other aspects of the improved service and any value of 'fraction broken' as a critical measure would disappear. It would certainly be foolish to devise, purely for assessment purposes, fragile test loads so that a measurable proportion of them still got broken. Subjective appraisal of a service is, indeed, invariably based on unconscious consideration of more than one factor, with each of which can be associated some criterion of satisfaction; this is certainly so in the case of speech links. The subject using the link ranks the criteria in order of importance for his general purpose—conversation in this case—and in making a judgment he has a pronounced tendency to take for granted those factors that have attained a sufficient level and to concentrate on the most important one still not fully satisfied. Thus the particular factor on which critical judgment is exercised is itself liable to change with the level of performance. This remains true even when there is only a single physical impairment variable, such as attenuation or signal/noise ratio, at work in a particular speech link.

Consider two separate links that have to be compared over a range of added line noise. For some values of noise valid assessments can probably be made using a measure such as 'proportion of read sentences correctly received' which is based essentially on one-way speech transmission rather than on conversational exchanges; this will be applied in the direction for which assessment is required. As the noise increases and the quality of transmission decreases a point will be reached where the assessment may become more realistic if it measures the effort needed by the listener in framing and successfully requesting repeats of fragments not understood; many different factors including reverse transmission, which previously did not matter, now come into play, and a different technique of measurement must be used. Unfortunately, the decreasing relevance of the first method as transmission conditions get poorer does not give rise to any increased difficulty in conducting the test itself, and the experimenter may therefore remain unaware that his technique of measurement is becoming irrelevant. The range of transmission quality over which some particular measure, say sentences received correctly, is applicable is by no means obvious and the determination of approximate boundaries to these ranges is an essential background to the work of assessment. Seriously misleading results can follow when they are crossed without proper precautions.

The difficulties spring from the peculiar nature of the speech signal itself and from the way two human beings co-operate with one another to exploit the properties of any transmission link in achieving their purpose of conducting conversation. The nature of the input information is ill defined, as are the channel capacity (since the physical extent of the channel itself is not obvious) and the output information. It is best to start with an elementary link, the one-way communication path shown in Fig. 1, where these difficulties are not present and where the



Fig. 1.—Elementary communication channel.

channel capacity, C bits per second, is definite. The source of information, A, has an information rate, H_i bits per second, assumed constant and uninfluenced by everything that follows. The sending apparatus, B, transforms the information signal to suit the channel; the transformation is often merely one of physical form, such as from literal or acoustical to electrical, but it may involve coding. The receiving apparatus, B′, performs a complementary function and yields a signal, whose information rate is H_o, to the receiver A′ who is assumed to require it in a similar form to that in which it left A. It is now well known[9] that H_o can, with optimum coding at B and decoding at B′, attain but never exceed C bits per second; H_o is also obviously limited to the input rate H_i. Thus H_o/H_i, the ratio of the output to the input information rate, is limited to unity or C/H_i (whichever is the less), and this measure of attainable efficiency is a function of channel capacity, as shown in Fig. 2.

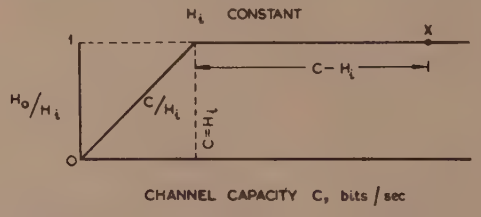


Fig. 2.—Theoretical communication rate of elementary channel.

Compare the case of telegraph transmission, where it is easy to measure performance, and of speech transmission, where it is not. The signal elements in the telegraph case are substantially independent of one another and must be transmitted virtually free from error; thus working conditions must be such that $H_o = H_i$. Practical telegraph channels are thus normally arranged to have a capacity, say C_x, such that the working point, X, lies on the horizontal part of the curve and the margin of excess capacity, $(C - H_i)$, acts as a reserve against deterioration of transmission conditions whose effect is to cause the value of C temporarily to wane. The margin may safely be made small in the case of line circuits, but may have to be large with, e.g., long-distance radio links subject to fading. However variable the transmission conditions, it is still possible to satisfy the user so long as C is not reduced to bring the working point onto

e slope. When this does happen errors of some kind become evitable; their effect can sometimes be partly avoided by rious means, but in general the sloping part of the characteristic of little interest for a telegraph link.

Now assume the same diagram, Fig. 1, to represent a one-way eech transmission system in which A is the speaker and A′ the tener. B and B′ are now the electro-acoustic transducers; in nventional systems there is no element of coding and decoding B and B′.

The full line in Fig. 3 represents, rather abstractly, the per-

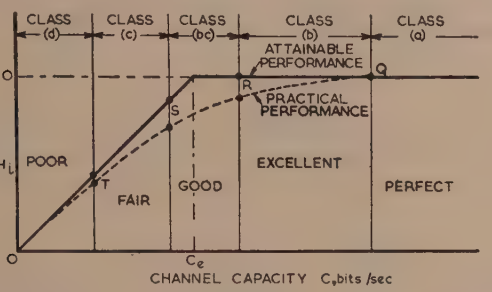


Fig. 3.—Performance of a speech link as a function of channel capacity.

rmance of a one-way speech path in terms of theoretically tainable values of H_o/H_i. The simplest way of estimating a ugh value for H_i, the information rate offered to the path, is regard speech and hearing as complementary techniques and take H_i as the information rate which the human ear can sorb and resolve. Estimates based on differences detectable the ear suggest a figure in the range of 5000–20000 bits/sec; r the sake of definiteness it will be taken as $C_e = 15000$ bits/sec. n exact value is quite unimportant and, indeed, even an accurate finition of what is wanted is difficult. Some aspects of this bject have been discussed by the authors elsewhere[21] and by uastler[25] in the related field of message information. The ed here is the relatively crude one of estimating roughly the ast channel capacity that has some prospect of conveying the eech in its entirety. A channel efficiently used, with B and providing any necessary coding, would yield approximately e performance shown; in particular, perfect transmission of eech with all the subtleties that the ear could appreciate would possible as soon as C reached the value C_e, after which the rformance curve would remain horizontal. Practical per-rmance for the one-way case is represented by the broken line d, except at the extremes, this is everywhere less than the eoretical (full line) if the channel is inefficiently used, as it is all physically realizable cases, especially if B and B′ are nple transducers. An example of a physical channel which s a capacity equal to C_e (15000 bits/sec) is one of pass-band 0–2800 c/s and signal/noise ratio 16 dB; but this, used with dinary transducers, offers far from perfect speech transmission. value of H_o/H_i of unity, corresponding to a condition in ich the output and input speech are indistinguishable by the r, needs with practical channels a vastly greater channel pacity, probably of the order of 3–5 times C_e, say 50000–000 bits/sec.

As the capacity of a typical practical channel is reduced from high value to something not far above C_e bits per second, a te is reached where the received signal becomes patently fective and the listener must increasingly use one or both of o techniques to maintain his continuity of understanding. e first involves the use of a return path and will be considered er. The other is the exercise of mental effort by the listener exploit knowledge of the probable nature of the speech signal. ost of the mental processes needed when the capacity exceeded C_e by a large margin (e.g. for conditions to the right of R in Fig. 3) were accomplished with at most a barely conscious listening effort in this sense. The combination of ear and brain has great ability to understand speech impaired in transmission, and consequently, to the listener, the fall in the information content of the received signal is gradual. Nevertheless, the effort to maintain full understanding of all that is said at the rate at which it is spoken and in the face of a decreasing amount of total information is an increasing mental strain and diverts mental energy that would otherwise be available for the conversation itself. Much speech communication is, however, conducted over conventional links whose channel capacity is below C_e.

There are thus *prima facie* three regions of interest:

(*a*) Capacities comfortably above, say, 50000 bits/sec, where signals received by A′ should be nearly indistinguishable from those leaving A.

(*b*) Capacities from about 50000 bits/sec down to about C_e corresponding more or less to the flat working region in the telegraph case.

(*c*) Capacities below C_e, i.e. below the limit where all the information of which the ear could make use is, or could be, transmitted. The talker can only be understood (in the absence of a return path to form a complete speech circuit over which the listener could request and receive repetitions) if the listener consciously diverts mental effort to deducing the signal from his store of knowledge and experience.

Five classes of transmission, (*a*), (*b*), (*bc*), (*c*) and (*d*), are shown in Fig. 3; the four boundaries between them have been marked on the curves as Q, R, S and T. Classes (*a*), (*b*) and (*c*) correspond roughly to the three regions of interest listed above. The transition from one class to another is far from clear-cut, and as this is particularly so from (*b*) to (*c*) and because this range of transition has a special significance which will appear later, the extra class (*bc*) has been inserted between them. The lowest region of interest has also been divided to yield two classes (*c*) and (*d*); the lower of these, corresponding to the most adverse conditions under which speech communication is practicable, is characterized by the need for appreciably increased time for the conduct of all conversational exchanges and consequently a slower rate of conveying ideas and information (see Section 3).

Though the whole Figure is only a crude representation of what is happening, it is important to realize that the range of links that are practical, in the sense of being neither absurdly expensive to provide nor impracticably poor for transmission, is a wide one and embraces all five regions. For example, the channel capacity of a music circuit,* which is about the cheapest that might transmit speech nearly indistinguishable from the original, is probably about 100000 bits/sec, while that of a line at roughly the lower limit of commercially acceptable telephone quality and including, as part of the link, the effects of reasonable acoustic noise at the receiving end, is well below 10000 bits/sec. Links used for tactical purposes in defence communications may, however, amply justify the use of still lower channel capacities. The smooth passage between these regions, even with the one-way link so far considered, obscures the change of conditions and the corresponding shift of criteria by which the listener judges the quality of the link.

(2.2) Effect of Return Path

When a return path is introduced, forming a circuit, the situation changes considerably. The effectiveness of an ordinary

* The term 'music circuit' is usually applied to special line links of 6 kc/s or more bandwidth and good signal/noise ratio, suitable for programmes for broadcasting and other special purposes. The link is not necessarily a circuit since it may be one-way only.

conversation depends a great deal on the establishment of linguistic contact and through it a measure of confidence and mental contact between the subjects. (The situation is again different when a talker is addressing several people, but these conditions are of secondary interest in telephony.) When the circuit is a very good one, giving conversation perhaps equivalent to that in noiseless conditions over a few feet of air, there is a tendency for the subject, while talking, to down-grade the quality of his speech in various ways, such as by reducing his vocal level, by the use of fragmentary and ill-formed sentences, poor articulation of sounds and mumbling of words, if he thinks the conditions will stand it. The talker knows, or soon finds out, that there is an equally good path for return speech and that consequently he will very quickly sense how the other subject is accommodating himself to the transmission. As the conversation proceeds, the talker reduces his effort on both the articulatory and mental aspects of his speech until the conversation is proceeding at a point corresponding to, say, somewhere between Q and R, where the attainable and practical one-way curves are substantially horizontal and have only slightly separated. The effect of introducing the return path is thus complicated; it causes variations of channel capacity to become confused with variations in information content of the signal from the talker and makes it practically impossible to study separately the behaviour of the subjects and the effect of varying the channel itself. It almost has the effect, as very careful thought will show, of bringing the effective two-way or circuit performance back to the track of the solid curve even though the broken curve accurately represents the one-way performance of each path separately.

A point of contrast with the telegraph case is also interesting; a good transmission channel there yielded a margin, referred to earlier as $(C - H_i)$, which was absorbed whenever the channel deteriorated. In the telephone case good transmission is exploited all the time by the subject to ease his personal talking burden in conducting the conversation.

Broadly then, however good the link, a practical conversation tends to be conducted at a point where further degradation due to the insertion of a channel of limited capacity is likely to pull the instantaneous conditions onto a more sloping part of the broken curve where conscious *listening* effort in the sense already discussed would have to be exerted; the effects of conversational exchange will, however, soon cause the other subject to realize this and he will up-grade his talking effort until an equilibrium is re-established somewhere between Q and R on the curve.

It is seen that the ordinary arts of conversation are likely to ensure a reasonable division, between the actions of talking and listening, of the small amounts of extra effort so far involved. With the full circuit, while neither subject ever enjoys, even with perfect transmission, quite such perfection of received speech quality as does the listener with the perfect one-way link (where he could not be exploited by the talker), the two subjects by exercise of sympathy and co-operation conduct a conversation with no more effort than if they were together in the same quiet room; to provide a realistic datum it is convenient to regard these conditions as requiring zero *conversational* effort. This condition proves to be resistant to slight further impairment in the transmission, owing to the 'feedback' effects round the circuit, and conditions may be imagined as moving more nearly along the solid than the broken curve, only departing from unity between R and S.

(2.3) One-Way and Both-Way Assessment Methods

Provided that the transmission conditions are somewhere in the first two classes, (*a*) and (*b*) in the above list where dotted and full curves have not separated far, only at most a slight listening effort would have been required even under one-way conditions, and the existence or otherwise of a reverse (A′–A) path is unimportant in assessing the value of the link for transmission from A to A′. Circuits composed of two such paths, one in each direction, may thus be fairly appraised by tests based on using them as one-way links. Two statements are true of the circuit consisting of two one-way paths, one in each direction, whose transmission is located at the lower limit of class (*b*):

(i) It is composed of the worst paths that require no appreciable listening effort to follow what is being said.

(ii) It is a comfortably better circuit than that which first requires appreciable conversational effort (as distinct from listening effort, which has meaning only in the one-way case) by the subjects using the link. A more precise meaning of the term 'conversational effort' will appear later.

If (i) is regarded as defining the lower edge of class (*b*), the onset of conversational effort is substantially below C_e, say at S. The class (*bc*) therefore corresponds to the need for listening effort under one-way conditions but no conversational effort when the link is two-way. This feature has been demonstrated experimentally and is discussed later.

When the circuit is below class (*b*) the passing of information becomes more and more a matter of co-operative effort between both subjects, and the division of effort between talking and listening becomes indistinct; both-way circuits are essential to assess these conditions. Fig. 4A depicts the conditions of a practical telephone conversation with two-way transmission. In the upper path, A is the point in the brain-speech transmission

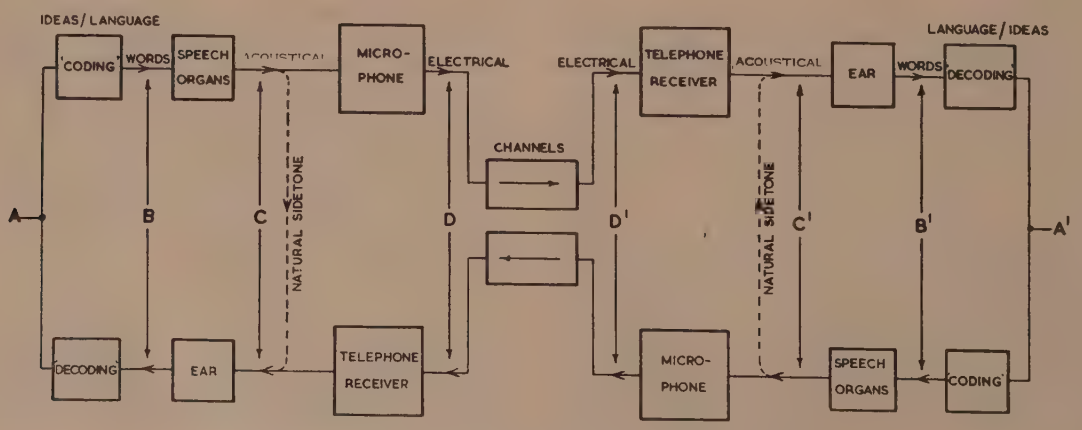


Fig. 4A.—Conditions of a practical telephone conversation.

ıain which is characterized by being the stage at which ideas re complete and in a form suitable for framing or 'coding' into ords. At stages B, C and D the words all exist though respecvely in the form of intended speech, acoustical signals and ectrical signals. The content here usually suffers little change, ıough its medium changes greatly, but noise and distortion of arious kinds can pare away some of the information content etween B and D. The speech sidetone path shown at C is an ssential feature of the transformation of intended speech into coustical signals and provides a feedback path for monitoring ıe emitted speech. Dramatic effects are produced when this ath is disturbed in certain ways.[12] After the transmission ıannel there are complementary stages, whose function is bvious, leading to A′ which is again the point at which ideas re handled.

For comparison, a high-quality reference system as used for ıe more elaborate transmission performance studies is shown in ig. 4B. The use of this type of reference speech link is discussed

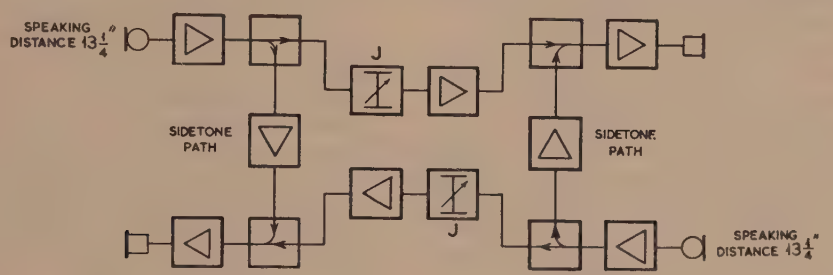


Fig. 4B.—High-quality reference speech link.

When attenuators J are set at 30 dB, mouth-to-ear loss is equal to metre-air-path conditions. Equipment is used in silence.

ter in the paper. With the specification of ambient conditions, ainly noise, at the two ends and addition of the two subjects emselves, the reference system is, in principle, indistinguishable om Fig. 4A.

The transmission paths for the opposite directions, also shown the Figures, have the corresponding stages and are self-xplanatory. At all stages between A and A′ in Fig. 4A the vo directions of transmission are everywhere clearly disnguished from one another, but outside these limits, i.e. at oints beyond the outer extremities of the Figure, they are at ach end merged in the process of thinking. The elements of iscussion and cross-questioning are inherent in communication ver a speech link and require the complete loop. For assessent purposes the omission of the return path is only admissible nder the restricted conditions discussed, for which a one-way nk will serve, and it is only for transmission between the points and A′ that information can be considered as, so to speak, owing round the circuit and so employing the arts of ordinary onversation.

The dilemma of a universal method for assessing speech nks is now evident. Assessment over a wide range of conitions is only possible if a complete circuit is used. A complete rcuit can only be achieved by considering communication etween points at or beyond A and A′; this in turn requires clusion in the link not only of the speech and hearing organs ut of those brain activities associated with thinking and with lea-language transformation. Thus to make the assessment eaningful we have now had to include in the link much that not itself part of the equipment but is part of the human user. hese human parts of the circuit are liable to be of both comaratively wide variability and unknown distribution. These re circumstances that make accurate knowledge of the perrmance of the equipment extremely difficult to acquire.

The only types of link that may safely be rated by one-way methods, e.g. between points C and C′, are thus:

(*a*) Links of high performance where conversational exchange, even with poor talkers, listeners and conditions of use, is unnecessary to elucidate anything that is being said.

(*b*) Links whose method of operation precludes return speech. These are relatively small classes of equipment.

(3) GENERAL METHOD OF ASSESSING TRANSMISSION PERFORMANCE

(3.1) Background of Assessment

The general method of assessment described below provides a set of practical techniques which fit closely into the framework described in Section 2.

Two different types of basic test are essential: one-way and two-way (or conversational). The first is based on listening to arbitrary non-conversational speech material, such as isolated words or sentences, and is conducted with no return speech path. It is safely applicable where the speech links are clearly within classes (*a*) or (*b*) in Fig. 3. The criterion for when it may be used is that inappreciably greater listening effort should be required of the subject than he would need in listening to the same (one-way) material under perfect conditions; in effect this means that he would not use a return path, even if available, so long as the material was of a one-way nature.

Attention should be drawn to a type of transmission degradation which affects talking and not listening, namely that likely to disturb the talker by its effect on natural speech sidetone, already mentioned. It is convenient to regard this as giving rise to talking effort the onset of which also moves the speech link from class (*b*) to class (*bc*) even though no listening effort is called for. Although measurement of talking effort is by no means simple, this effect can usually be disentangled from the factors which affect listening and conversation and, in any case, it rarely occurs in serious form. For this reason and to avoid impeding the general argument, talking effort will merely be noted in the relevant places.

The component tests of a complete general assessment method must, in one way or another, detect (and possibly measure) the following:

For one-way conditions	(i) The extent to which the reproduced speech can be distinguished from the original or what would be received over a direct air path.
	(ii) The use of listening (or talking) effort.
For two-way or conversational conditions	(iii) The use of conversational effort.
	(iv) Whether extra time is being taken up on account of transmission difficulties.

On each one of these can be based a criterion for defining a boundary between different classes of conditions as the speech

link becomes more degraded; the four boundaries separate substantially the five classes marked in Fig. 3. Of these, however, the boundary between the uppermost two classes depends only on the rather aesthetic question of whether the speech is distinguishable from the original, and is clearly unimportant for telephony purposes; no work has been devoted to devising tests to locate it. It is, however, retained in the assessment framework (see Table 1).

In assessment, therefore, we have to (i) define and detect the class boundaries; (ii) make sure in every specific case that we use experimental techniques appropriate to the level of performance; (iii) devise a method of measurement or scoring within each class. These objectives are discussed below.

(3.2) Delineation of Classes and Detection of Boundaries

Many of the factors which influence transmission performance, such as the listening ability of the users and the difficulty of the spoken material, are themselves of wide distribution, and hence a criterion defined in terms such as 'just zero mental effort' presents difficulties as a basis for actual tests. Each criterion is therefore taken as satisfied when 95% of the transmitted material can be regarded as acceptable by that criterion to the median member of the population of ordinary men and women. It is thus implied that the various tests must employ not a special trained crew but a fair sample of normal persons.

Two classes of transmission performance have now been defined by reference to conditions with no return-path transmission: that in which more than 95% of the received material is indistinguishable from the original, and that in which this is not true but more than 95% of the material demands no more effort of the listener than he would use while listening to speech transmitted by a perfect circuit.

If the return speech path were in operation the talking behaviour of the subject would be affected, since it would be influenced by his judgment of how much listening effort the other subject was being called on to exercise. This he would subconsciously do his best to adjust to the amount experienced in ordinary direct conversation. If the transmission performance is in either of the two high classes already defined, the subject who is talking will sense, as soon as the reverse path is opened and two-way conversation becomes possible, that he has something in hand and can consequently safely degrade his talking effort before the other has to use appreciable listening effort. If the automatic regulating effect that this model of the talker's behaviour implies is accurate, the conversational performance of a link in its normal two-way use is substantially independent of the quality of its transmission, so long as it lies somewhere in the top two classes. Over this range the talking effort will decrease (assuming that natural speech-sidetone conditions obtain) rather than increase when the return path is added; in fact the lowest limit of these classes, (*a*) and (*b*), will be the last point at which the link provides conditions that are substantially those of ordinary direct conversation. This is in exact accord with experience; over an indefinite range of channel capacity from, say, 15 000 bits/sec upwards there is virtually no change in the value of a link for conversation.

At the bottom of class (*b*) the conditions are such that the talking effort will be just unaffected by the connection of the return path. For all transmission worse than this the subject will increase his talking effort as soon as two-way conversation is possible, as he will sense that the other is having more listening difficulty than in direct conversation.

The delineations of the first two classes were based on one-way conditions; the lower classes obviously must be discussed entirely in terms of both-way conditions. It is found that, if the both-way circuit which at the lower limit of (*b*) could just be connected without effect on effort is now further degraded slightly, communication remains possible without appreciable continuing conversational effort. As will be shown below, circuits that fail by the first two criteria are usually, over a further range of degradation, still fully satisfactory for conversation as soon as the subjects have established effective mental contact. This is only partly the direct result of both subjects getting used to one another; it is also due to the self-regulating effect of each subject's talking: he will, for example, articulate more carefully now in an unconscious attempt to bring the conversation back to direct air-path conditions. The third criterion, which defines the lower limit of this class [termed (*bc*)], is that there should be no appreciable conversational effort on the both-way link.

Once we pass beyond this stage and admit conversational effort, a further definite range of degradation [class (*c*)] is possible in which the artifices and subtleties of conversation have to be increasingly employed. When there is little conversational effort the rate at which requests for repetitions are made is negligible—no greater than in ordinary conversation—but the need for repetition and for paraphrasing and circumlocution by the talker becomes more frequent as degradation increases; also, the speech tends to become slower and it takes longer to piece together what has been said and to respond.

These effects are slight at first, and a surprising amount of further degradation is possible before the actual time taken to convey a given amount of information increases. The criterion defining the boundary between classes (*c*) and (*d*) is therefore that the rate of conveying information by ordinary conversational methods (termed the 'message rate') should be at least 95% of the rate for direct conversation.

Class (*d*) covers conditions failing to pass even this last criterion. Most subjects describe such circuits as poor, but until the time needed has increased considerably they are not unusable. With a poor circuit the extra time taken by the subjects to get their information across is a fair measure of the unsatisfactoriness of the circuit. There is no definite limit to the poverty of circuit that can be pressed into service, but the time taken increases rapidly beyond a certain point, and even the most determined subjects are unlikely to be able to make much use of circuits requiring more than two or three times the normal conversational time.

The distinctions between the classes are summarized in Table 1.

(3.3) Experimental Techniques

Suitable experimental techniques for the general assessment method can best be illustrated by means of tests on a specific speech link. The link here described is formed from the circuit shown in Fig. 4B supplemented by circuit noise injected into each listening channel. The noise is of continuous-spectrum type shaped in frequency spectrum to conform to that of average speech. A convenient feature of this noise is that the ratio of mean speech power to noise power measured in any narrow frequency band is independent of the mid-frequency of that band. A constant-volume amplifier having a time to operate* of several seconds is incorporated in each sending channel to ensure that the ratio of mean speech power to noise power is independent of the speaking level. The speech level at the ear is adjusted to correspond to that obtained under average conversation conditions in the absence of any ambient noise; this speech level is maintained independently of the injected noise level at the ear, which is varied as an adjustable parameter. The physical performance of the link is thus specifiable in terms of the ratio of mean speech power (averaged over the period during which

* 'Operating time' for equipment such as a constant-volume amplifier is analogous to time-constant for a simple circuit, but with signal levels replacing direct currents or voltages. Exact definition is complicated; the subject is discussed in Post Office Research Report No. 13101. (Restricted availability.)

Table 1

CLASS CRITERIA FOR TRANSMISSION PERFORMANCE OF SPEECH LINKS

Class	One-way speaking		Two-way conversation		Rough description of transmission performance (see Note 2)
	Extent to which link is distinguishable from noiseless direct air path	Talking/listening effort (see Note 1)	Conversational effort	Message rate	
(1)	(2)	(3)	(4)	(5)	(6)
(*a*)	Barely or not at all (see Note 3)	None	None	Normal	Perfect
(*b*)	Appreciably	Slight or none (see Note 3)	None	Normal	Excellent
(*bc*)	Completely	Appreciable	Slight or none (see Note 3)	Normal	Good
(*c*)	Completely	Inapplicable; one-way tests of little value	Appreciable	Slight or no reduction (see Note 3)	Fair
(*d*)	Completely		Considerable	Appreciable reduction (more than 5%)	Poor or useless

Note 1.—In most practical cases the onset of appreciable listening effort marks the boundary between classes (*b*) and (*bc*). Talking effort would arise when severe speech sidetone or echo were present while talking.

Note 2.—The terms in column (6) are loose descriptions only and do not form part of the description of the classes. They are roughly in accord with a comment which might be made by an average man using the link for a brief conversation under conditions free from physical stress. They must not be taken as a guide to user evaluations.

Note 3.—'Slight or none' is taken to mean that less than 5% of material requires, on the part of an average man, the type of effort to which the column refers. Similarly, slight or no reduction in message rate means up to 5% reduction. A similar meaning is attached to 'barely' in column (2).

Note 4.—It will be seen that all class boundaries are described in terms of the onset of appreciable degradation of performance under one of four criteria [columns (2), (3), (4) and (5)]. The class boundaries described in this Table have the same significance as those marked in Fig. 3.

eech sounds are being uttered) to the mean (unweighted) noise wer.*

When the circuit noise is adjusted to any value that renders it udible, the speech link provides virtually perfect transmission d so qualifies for class (*a*). By increasing the noise level the k can be gradually degraded until the speech/noise ratio ally becomes so unfavourable that no speech signals can be ected; speech communication becomes impossible, however, ll before this occurs.

As mentioned earlier, the transition from class (*a*) to class (*b*) of little interest in telephony. The other three boundaries uire measurements of effort and message rate. Effort can be asured (or at least the need for it detected) either by deter-ning the proportion of material transmitted without effort or use of a subjective assessment scale whose categories are tinguished by descriptions indicating increasing need for ort; both techniques prove practicable both for listening ort [giving the (*b*)/(*bc*) class boundary] and for conversational ort [(*bc*)/(*c*) class boundary]. Message rate is measured by ing the performance of a definite task which subjects must nplete with the aid of the speech link.

The particular type of degradation considered here does not olve the use of talking effort, at least until the noise is well ond the level required to call for listening effort, and talking ort need not therefore be discussed.

The upper curves in Fig. 5 show results of two different kinds assessment related to listening effort and applicable under one-way conditions which serve to define, along the scale of degradation, the boundary between classes (*b*) and (*bc*). The percentage of sentences understood without listening effort was obtained by means of recorded lists of sentences reproduced to a listener who was required to indicate, after hearing each sentence, whether he had understood its meaning without listening effort, i.e. immediately and without any desire to ask for a repetition. The proportion of sentences so understood is recorded. This method is similar in principle to the immediate-appreciation method devised by Grinsted,[4] but there are two important qualifications. First, subjects for the present purpose must consist of a sample of ordinary untrained and unpractised persons; this implies that a team of operators cannot be used and that the subjects may not be employed more frequently than, say, one experiment (lasting a total of a few hours spread over several days) every six months. The second qualification is that the test takes place with no preliminary both-way conversation over the speech link. Lists of only 25 sentences are used and the amount of degradation in the speech link is changed after each list; sessions of listening are also restricted to 20 or 30 min.

Where most of the sentences are understood without listening effort the link falls in class (*b*), or class (*a*), and so the setting corresponding to a score of 95% is taken as the boundary between classes (*b*) and (*bc*).

Clearly, it is difficult to determine a setting corresponding to such a high percentage with very great precision owing to the small slope of the curve in this region of score; fortunately the listening-effort assessment-scale method yields more precise results and leads to the same value for the boundary. This method depends upon the subject indicating his opinion on the following assessment scale after listening to a group of five

* This method of expressing speech/noise ratio must not be confused with al/noise ratios quoted for trunk telephone lines, which are often expressed as the of maximum permissible sinusoidal power to weighted noise power; such figures be at least 25 dB more favourable than any quoted in this example to cater for eaks of speech waveforms (say 12 dB greater than the mean power) and the wide e of speech levels in the absence of any constant-volume device. Speech levels lly have a standard deviation of about 5 dB.

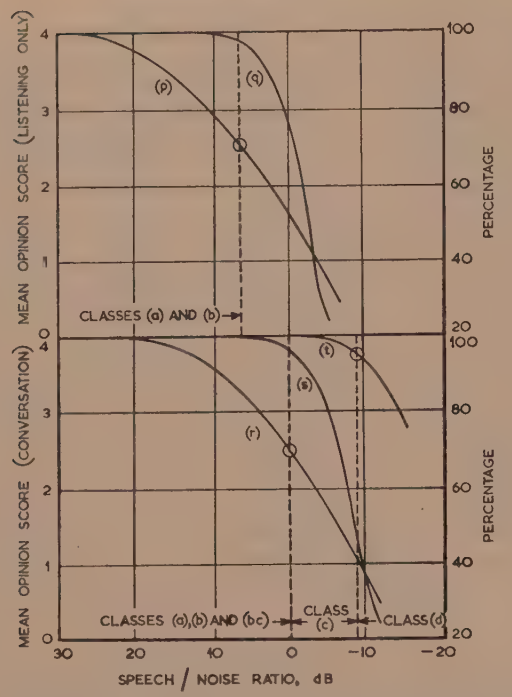


Fig. 5.—Transmission-performance measures which define the inter-class boundaries: high-quality reference speech link with injected noise.

(*p*) Mean opinion score (listening effort scale).
(*q*) Percentage of sentences understood without effort (listening only).
(*r*) Mean opinion score after conversation (effort scale).
(*s*) Percentage of sentences understood without effort after conversation.
(*t*) Percentage message-rate efficiency.
Speech level is expressed as r.m.s. value. Speech voltage is about 6 dB higher than r.m.s. value.
Noise (continuous spectrum shaped to conform to that of speech) is measured as r.m.s. value unweighted.

sentences reproduced from a recording. A test consists of reproducing ten groups of sentences over ten different settings of the link, used in a random sequence, and collecting the ten respective opinions; these can conveniently be signalled by keys to avoid the listener having to write or speak.

The categories of the scale are described as follows:

Opinion Assessment Scale based on Effort.

A—Complete relaxation possible: no effort required.
B—Attention necessary: no appreciable effort required.
C—Moderate effort required.
D—Considerable effort required.
E—No meaning understood with any feasible effort.

The results for a number of talkers (used in preparing the recordings) and listeners arranged in accordance with a suitable experimental design[13] are scored numerically as 0 for *E*, 1 for *D*, 2 for *C*, 3 for *B* and 4 for *A*. The mean values are plotted in the upper part of Fig. 5. The mid-point between opinions *B* and *C* may reasonably be taken as the effort threshold, and so a mean score of 2·5 may be taken as marking the boundary between classes (*b*) and (*bc*). Both methods thus yield +6 dB speech/noise ratio as the position of this boundary.

In the example chosen the received speech level is unaffected by the speech level from the talker or by the setting of the link; where this is not so the speech material must be reproduced for one-way tests at a received level corresponding to that obtained when a median person is speaking over the link without the return path.

As soon as the degradation takes the link beyond class (*b*) into class (*bc*) purely one-way techniques must be replaced by conversational methods. The subjects are now employed i pairs, one at each end of the both-way link, and conversation q a vigorous, purposeful nature is generated by presenting ther with a situation in which certain information must be exchange between them to solve a pictorial puzzle. Random shapes[16] ca be used in such a way that they must be matched against verbal description by the other subject. The task is designed t occupy two or three minutes over a good link; on completio the subjects have achieved as much mental contact as the con ditions will permit and are asked to express their opinions of th link on the same effort scale as was used for one-way conditions In this case they will, however, be basing their opinions on con versational experience and not on mere listening experience Scoring as before yields the curve in the lower part of Fig. marked 'mean opinion score after conversation'. Again mean score of 2·5 marks the onset of effort, in this case conversa tional, and, as can be seen, 0 dB marks the boundary betwee classes (*bc*) and (*c*).

If, immediately following the conversation, the same pair o subjects is given lists of sentences for each to read to the other the listener, by saying 'yes' or 'no', indicating to the reade whether he had understood the meaning of each sentenc without mental effort, scores can again be plotted of percentage of sentences understood without effort after conversation. I this case higher scores are recorded than in the purely listenin situation, corresponding to a displacement of the setting whic yields 95% score from +6 dB to 0 dB. Thus the 'mean opinio score after conversation' and the percentage of sentences under stood without effort after conversation agree in defining th point of onset of conversational effort.

The puzzle-solving stage of the conversational tests is time and the time required is compared with that in a separat test for the same pair of subjects to solve a similar puzzl over a class (*a*) or (*b*) speech link. The ratio of the latte time to the time taken over the link has come to be terme message-rate efficiency. This commences to fall below 95% a speech/noise ratios substantially worse than that which mark the onset of conversational effort; for the present example it i at about −9 dB. By argument according to Section 3 it is t be expected that increasing conversational effort would at firs succeed in overcoming the need for additional time, but wit increasing degradation of the circuit the maximum conversationa effort of which the subjects are capable will eventually b exerted. At this stage any further degradation results i increasingly lengthy discussion of repetitions and the adoptio of circumlocutory phrasing; consequently the message-rat efficiency falls fairly suddenly beyond the (*c*)/(*d*) class boundary which is therefore quite well defined.

Although conduct of any ordinary conversation over a class (*d* link is at best a serious strain on the subjects, there is a wid range of circumstances in which it is impracticable to provid anything better. When degradation has reduced the perfor mance to this final stage in which an indefinitely long time ma be needed to convey any message content at all, the message rat can approach zero. This takes us beyond the worst links tha could have any value for communications, and message-rat efficiency thus remains a satisfactory assessment measure for th poorest circuits that are of any practical value.

(3.4) Practical Application

In its most general form the assessment method describe above consists in applying various tests designed to expose th class in which the link lies by the criteria set out in Table 1 this in itself provides a sound but coarse assessment. Actua figures obtained with the test appropriate to the class, e.g. 'mea opinion scores', provide a finer measure such as is commonl

quired for the uses listed in Section 1. The means of applying e general method to problems concerned with transmission erformance rating will now be illustrated. This corresponds (*b*) of the list of uses, but the same type of assessment informa- on is also applicable for (*c*), (*d*) and (*e*).

The problem considered here is that of rating a number of ommercial speech links relative to a common Standard. A ries of assessment measures as described in Section 3.3 is btained on each of the links under consideration, including the tandard, over a range of settings of each. These serve to ark the inter-class boundaries on each rating scale. Super- osition of these points on the scale gives, by the amount of splacement necessary, the ratings at the levels of performance efined by the respective boundaries. Fig. 6 shows such a

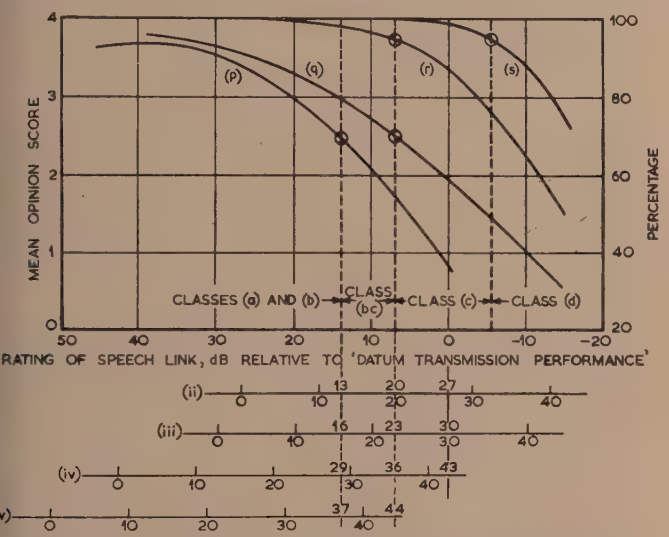


Fig. 6.—Transmission-performance measures which define the inter-class boundaries: British Post Office subscribers' sets.

(*p*) Listening mean opinion score (effort scale).
(*q*) Mean opinion score after conversation (effort scale).
(*r*) Percentage of sentences understood without effort after conversation.
(*s*) Percentage message-rate efficiency.

Supplementary scales represent loss between terminal exchanges for

(ii) earlier type sets, limiting subscribers' lines, 60 dB room noise;
(iii) later type sets, limiting subscribers' lines, 60 dB room noise;
(iv) later type sets, average subscribers' lines, 50 dB room noise;
(v) later type sets, zero subscribers' lines, no room noise.

mily of results plotted as a function of rating; 0-dB rating rresponds to the transmission-performance datum, explained detail elsewhere.[28] It will be seen that the boundary between asses (*b*) and (*bc*) occurs at a rating of +14 dB. Scales (ii), i) and (iv) in Fig. 6 are of loss between terminal exchanges for e three speech links used to obtain these curves;* e.g. the)/(*bc*) class boundary, corresponding to scale (iv), occurs at a ss setting of 29 dB; hence the rating of this link (with no loss) +14 + 29 = +43 dB. Scale (v) has been included to illustrate ie end of the enormous range of ratings to be considered in a ublic network, and it will be seen that this extends into the gion where the opinion score commences to fall owing to cessive loudness of the received signal.

The position of the (*c*)/(*d*) class boundary cannot be deter- ined at all precisely on these speech links because the poor nsmission in this region becomes increasingly compensated an increase in talking level which correspondingly raises the level at the listener's ear. Such poor links occur only rarely, of course, on a public telephone network.

The equipment and conditions considered in the above examples cover only a narrow range of type of distortion; in these circumstances a rating for one setting of junction loss can be deduced from that measured at a different setting merely by subtracting the difference in junction-loss settings.

While the curves relating to listening effort are being obtained the speech level reproduced from the talking end must correspond to that for average one-way speech, i.e. the effects of the reverse direction of transmission must be excluded. This speech level can conveniently be measured by taking a sample of ordinary subjects and asking them to read test sentences while holding the handset naturally or, in the case of the high-quality system, speaking at the correct distance from the microphone. The speech level reproduced electrically on the circuit of a commercial speech link depends, of course, not only on the vocal output of the talker but also upon the manner in which he holds his hand- set; for this reason it is not sufficient to measure the electrical speech level under formal conditions such as apply in trained- crew articulation tests,[22] where 'reference vocal level' and a conventional speaking distance and position are employed. It is possible in many instances to determine a relation between actual conditions of use and these fixed conditions that may safely be used for one particular type of handset.

Under conversational conditions the speech level will be adjusted automatically by the subjects according to the per- formance of the link, higher values being used as it becomes more degraded[20] and lower values when better conditions obtain than those corresponding to the (*b*)/(*bc*) class boundary owing to the talker taking advantage of the situation. Conversational speech level is also affected by other factors, including the level of sidetone, the room noise level and the degree of interest in the conversation.

So far, the results of transmission-performance measures as described in Section 3.3 and illustrated in Figs. 5 and 6 have been expressed as mean scores averaged over all members of the sample of subjects used. When a reasonably large sample is available it becomes possible to present the results in greater detail and to expose the dispersion of results between the different subjects. Some examples are shown in Fig. 7. Curve (*p*) is of the percentage of subjects who, during the tests leading to the curve of 'listening mean opinion score' given in Fig. 6, expressed opinions *A* or *B*, i.e. needed to use no appreciable listening effort. Curve (*q*) of Fig. 7 gives corresponding information for the

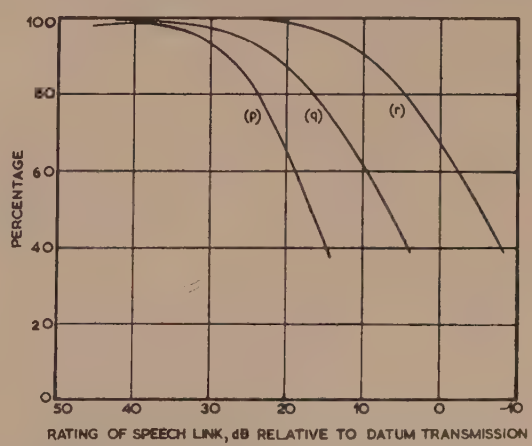


Fig. 7.—Percentage of subjects able to satisfy certain criteria as a function of rating relative to datum transmission.

(*p*) Percentage of subjects who require to make no listening effort.
(*q*) Percentage of subjects who require to make no conversational effort.
(*r*) Percentage of subjects who require to make no more than moderate conversational effort.

* In Fig. 6 'limiting subscribers' lines' means the most unfavourable lines allowable h the respective telephone sets; specifically 450 ohms 10 lb cable and 660 ohms b cable apply respectively for sets of earlier and later type used with 50-volt + 200-ohm feeding bridges. Room noise levels are expressed relative to a datum 0·0002 dyne/cm² at 1 000 c/s, an American sound-level meter being used; 60 dB resents a rather loud level.

conversational situation ('mean opinion score' after conversation). It would, in principle, be possible to present also the percentages for opinions '*A*', '*A* or *B* or *C*', and '*A* or *B* or *C* or *D*'. In practice, however, it is rarely possible to accumulate enough data for the two extreme cases, but curve (*r*) of Fig. 7 gives the results for opinions '*A* or *B* or *C*', i.e. the percentage of subjects who required to use no more than moderate conversational effort.

Data in the form of cumulative percentages of some quantal response are frequently met in biological assay, and probit analysis[14] is a powerful method for analysing such data. This consists in estimating, by the maximum-likelihood method, the values of location and slope parameters of the S-shaped Gaussian probability integral curve which best fits the observations. These parameters correspond to the mean (which equals the median for a Gaussian distribution) and standard deviation of a distribution of tolerance in the population of subjects. The word 'tolerance' in this sense [e.g. curve (*q*) of Fig. 7] is used to mean the minimum rating needed to ensure that the individual subject exerted no conversational effort. This technique can be used, without violence to the data, for ratings comfortably clear of that at which the percentages reach a maximum and start to decrease (e.g. lower than +30 dB in Fig. 7). A standard deviation of about 12 dB is commonly found for tolerance distributions of the kind discussed here.

Results in this form enable informative performance measures to be associated with the ratings for a statistical distribution of circuits and conditions; this corresponds to the problem arising in planning a large telephone network. As seen from Fig. 6, an enormous range of ratings is involved if all possible subscribers' lines and levels of room noise are to be considered.[28]

(3.5) Use of Reference Systems

A reference speech link, or reference system, is a laboratory speech link which can be specified physically in a simple manner so that it can be reproduced at different times and places and always have the same performance. Such reference systems are used for international telephony but, as discussed elsewhere,[28] there is also a need for an ultimate reference system based on transmission conditions over an air path.

Reference systems are used in much the same way as Standards (see Section 1) to determine the performance of unknown links; i.e. assessment scores are normally not used directly but for making comparisons leading to ratings in terms of the scale of degradation or impairment incorporated in the reference system. The setting of a generally available reference system thus provides a convenient means whereby the performance of equipment may be recorded, in terms of more or less permanent meaning, and hence held for comparison with that of any other equipment. The use of reference systems to obtain ratings is, however, more generally cumbersome and less precise than the use of ordinary Standards, one essential feature of which is similarity in transmission quality with the link being rated. Reference systems which aim at a measure of universality cannot do other than provide high-quality transmission, and are relatively unattractive as well as unnecessary for day-to-day use in rating.

The usefulness of a particular reference system is much enhanced when experience has been gained in its use, especially when assessments have been carried out on it over the whole usable range of its scale of impairment and the inter-class boundaries located on the scale. The results of a small programme of measurements using the general assessment techniques explained above are shown in Fig. 8 for an actual reference system based on metre-air-path conditions,[11] i.e. one whose 'variable impairment' has somewhere in its range a setting that yields transmission virtually the same as that of the ultimate

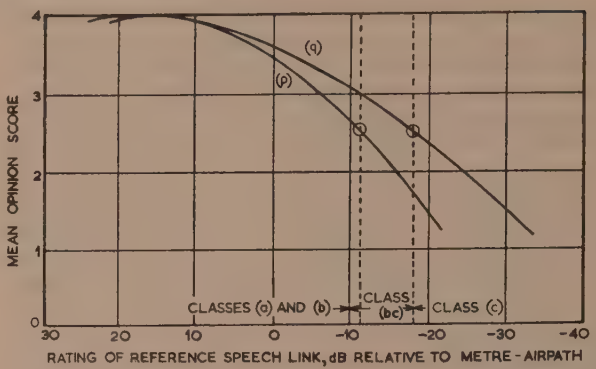


Fig. 8.—Transmission-performance measures which define the inter-class boundaries: both-way reference system based on metre-air-path conditions.

(*p*) Listening mean opinion score (effort scale).
(*q*) Mean opinion score after conversation (effort scale).

reference system defined in Section 8. 0 dB on the scale of degradation, which is one of loss, represents in this case the metre-air-path sensitivity. Listening-mean-opinion scores are plotted and the (*b*)/(*bc*) class boundary located at a setting of −11 dB. The (*bc*)/(*c*) class boundary located by means of the curve of mean-opinion score after conversation is at −18 dB. The other boundaries have not been determined.

Results in the form shown in Fig. 8, together with corresponding results for an unknown speech link, enable the latter to be rated in terms of the reference system. If the unknown contains commercial telephone sets it will usually be found that, unlike the situation mentioned in Section 3.4, their ratings in terms of the high-quality reference system are not independent of the performance at which they are measured. This will be seen if the corresponding curves of Figs. 6 and 8 are compared. Ratings in terms of such a reference system are, as has been seen, best reserved for situations where a Standard cannot be used.

The adjustable impairment in a reference system need not always take the form of loss. When the listening level is chosen by the subject, loss is clearly an inappropriate degradation, and added noise, a variable band-pass filter or a form of adjustable waveform distortion[18] has at various times proved convenient.

If a reference system is required for use where it will never need to be compared with any very good speech links, high-quality transmission is unnecessary; it may sometimes be advantageous to use a reference system only just better than the best unknown likely to be encountered by it. This, however, makes exact and simple specification less easy. It will be realized that the requirements of a reference system and a Standard usually tend to be rather at variance (see Section 8).

(4) OTHER ASSESSMENT METHODS

The assessment methods with which the paper is mainly concerned are of a rather general type, and their essential merit lies in the fact that they closely resemble the more important aspects of ordinary conversation. Most other methods are of more specialized application although many are very extensively used; they can conveniently be reviewed in order of increasing generality.

The transmission of a speech link can, of course, be measured purely instrumentally in terms of sound pressures, but, except perhaps in the case of very good (so-called high-quality) links, such results are, in the present state of the art, extremely difficult to interpret. Some progress towards this has been made by evolving various techniques to enable the articulation ratings of certain kinds of speech links to be calculated from instrumental

und-pressure measurements,[26] but the usefulness of this is, in y case, restricted to the field of application of articulation data. Calculation techniques such as these may also be based on bjective measurements of electro-acoustic sensitivity instead of measurements of sound pressure,[27] but the final objective ll remains that of ratings based merely on articulation methods. The next stage is the use of the minimum utterable speech ements as material to be transmitted, namely nonsense syllables 'logatoms'. This corresponds to the classical articulation test hich, when a trained crew is used, can be refined to give very ecise results. The procedure used in the British Post Office is scribed by de Wardt,[15] while that standardized in the C.I.T.T.* Laboratory in Geneva is described in C.C.I.F. erature.[22] Both these establishments use trained crews, hough the degree of proficiency of the C.C.I.T.T. crew is obably lower because its members do not possess a common tive language. The logatom material differs profoundly in the o cases; the British is based on the frequencies of occurrence phonetic elements, including frequencies of digraphs (permutaons of phonetic elements taken two at a time), in spoken nversational English,[17] while the C.C.I.T.T. Laboratory uses peranto logatoms containing most of the sounds, including mparatively rare and complicated phonetic combinations, und in written Esperanto.[22] All the Esperanto sounds are cluded with equal frequency irrespective of their actual degree rarity. Because of these differences the two establishments eld very different assessment scores for the same speech link, d serious differences are also found even in ratings of certain mmercial speech links relative to the same high-quality ference system. The same rather elaborate experimental signs and statistical techniques for analysing articulation sults[22] are, however, used in both establishments; they are sed on those originated by the British Post Office.

Words instead of logatoms may be used in articulation tests. e words may be monosyllabic or multisyllabic and may or y not be balanced phonetically.[10]

The use of carefully conceived experimental designs, however, w makes it possible to conduct articulation tests using trained subjects each of whom listens to only a very small mber of speech elements, so as to avoid any appreciable actice. Logatoms or words may be used, but, in spite of the gher scores when words are the material, similar ratings are in ese cases obtained with logatoms and with words. Use of dinary persons instead of a trained crew opens up the possiity of distinguishing between the purely listening situation d that which arises after vigorous both-way conversation. g. 9 shows English monosyllabic word-articulation results tained with untrained users, both before and after conversan, and with a trained crew. The results refer to the high-ality reference system with injected circuit noise. Fig. 10 ows results for a commercial link with untrained users after nversation, and with a trained crew. The latter, which were tained at a constant vocal level and under controlled talking nditions, are also shown after they have been corrected to take count of the increase in talking level (and hence speech ltages) as the performance deteriorates. The trained-crew ults in Fig. 9 require no correction because a constant-volume plifier was used. For untrained users the word-articulation ores at the corresponding inter-class boundaries lie between and 90% while the trained listeners still yield almost 100%. rthermore, it is clear that the trained-crew results are useful ly for discriminating between speech links which are already bad that ordinary subjects would have considerable difficulty conversing.

* The C.C.I.T.T. is a combination of the two former organizations, C.C.I.T. and .I.F. The merger took place in 1956.

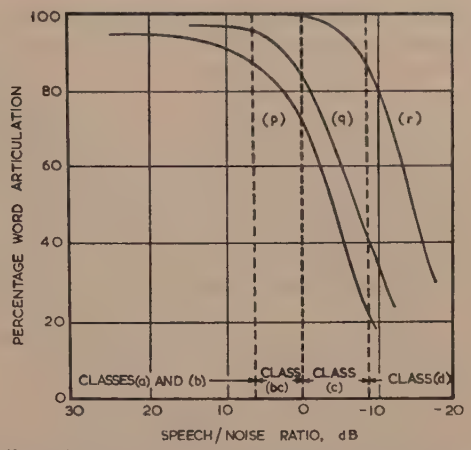


Fig. 9.—Effect of various methods on word articulation: high-quality reference speech link with injected noise as in Fig. 5.

(*p*) Normal users, listening only.
(*q*) Normal users, after conversation.
(*r*) Trained crew.

Articulation-like tests may be conducted with sentences as the material transmitted; scoring is sometimes in terms of the complete sentence written down or repeated by the listener and sometimes by counting the number of key words correctly received[10] (i.e. partial scores may be awarded for material not completely correct).

The essence of a sentence is its meaning rather than its separate words, and this idea is exploited in the immediate-appreciation method[4, 5, 6, 23] originated by Grinsted; sentences are spoken, but the listener indicates only whether or not he has understood the meaning immediately without any need for thought. As originally used the method involved using a team of subjects who became skilled and consequently yielded rather high scores, but, as explained in Section 3.3, it can be adapted for the purposes of indicating the onset of effort.

Meaningful material may also be used in a question-and-answer technique[1, 19] in which the listener is asked simple

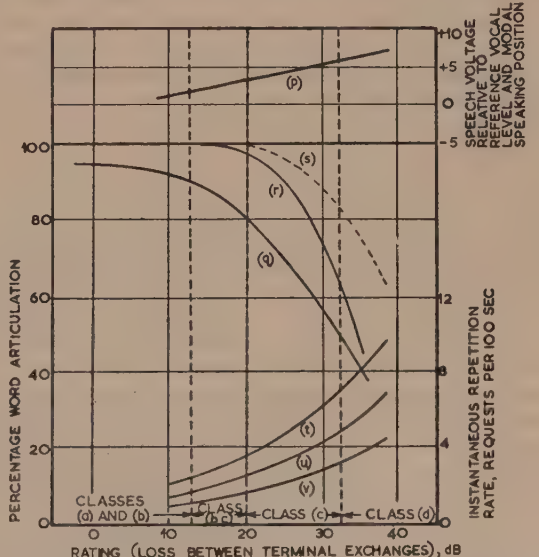


Fig. 10.—Word articulation and repetition rate as examples of other assessment measures: speech link as in Fig. 6, scale (ii).

(*p*) Relative speech voltage at sending end.
(*q*) Word articulation, normal users, after conversation.
(*r*) Word articulation, trained crew, fixed vocal level and speaking position.
(*s*) Word articulation, trained crew, corrected for sending-end speech voltage.
(*t*) Instantaneous repetition rate, initial.
(*u*) Instantaneous repetition rate, after 30 sec.
(*v*) Instantaneous repetition rate, final.

questions whose answers he is sure to know, such as 'How many days make one week?' Such material does not lead to any appreciable conversational exchange and so is not applicable to the purposes given in Section 3.3.

The most successful material for conversational studies is that which can be presented to the subjects without appreciable use of words, written or spoken. A system has been evolved by the authors for generating random shapes forming elements of a puzzle which may be solved by co-operation between the two subjects over the speech link. This forms the basis of the conversational-opinion and message-rate tests discussed in Section 3.3.

During 1931–38 a great deal of effort was put into building up a system of transmission-performance ratings based on observing incidence of requests for repetitions during actual telephone traffic. The method of observation is known as 'repetition-rate' and the system of rating based on it became known in the United States as 'effective transmission'.[2, 3] Repetition rate requires detailed monitoring of calls and obviously cannot be done on the public network itself for reasons of privacy. The tests made in the United States by the American Telegraph and Telephone Company and in the United Kingdom by the British Post Office were performed on private tie lines between large sections of the respective organizations situated in different places. Rating based on repetition rate was favoured by the C.C.I.T.T. at one time for international purposes. The position as it appeared in 1948 was commented on by Pocock.[7] It has, however, gradually become clear that repetition rate is not a suitable basis for rating for international purposes. Intensive statistical analysis of the Post Office results revealed[8] that the instantaneous repetition rate was not constant throughout a conversation but fell approximately exponentially with time; this was subsequently verified more directly by making repetition observations during the course of laboratory message-rate tests, and these results are included in Fig. 10. Instantaneous repetition rate is shown as a function of the rating, while time, measured from the commencement of the conversation, is used as a parameter. The rate of occurrence falls exponentially according to a curve having a time-constant of 40 sec. It is interesting to note that a slight rise in initial instantaneous repetition rate has occurred at the (*b*)/(*bc*) class boundary and a similar small increase in the final instantaneous repetition rate at the (*bc*)/(*c*) class boundary.

Although attractive, the method does not lend itself to rating purposes, but it might, perhaps, be useful in principle for contributing to a user evaluation.

(5) CONCLUSIONS

Experience has shown that it is not very difficult, particularly with the help of modern methods of experimentation, to devise assessment methods whose results are relatively precisely repeatable within a given laboratory. The difficulties lie in two further directions: first, in providing a method that will yield equally realistic results over a wide variety of speech links, and secondly, in securing a greater measure of repeatability between results obtained in different laboratories. Apart from lack of generality, existing methods have sometimes been found in practice to yield unrealistic answers to apparently quite simple specific problems.

Lack of a way round these difficulties has hindered progress enormously; the paper describes a general assessment method whose basis is closely related to the varying techniques of conducting conversation over links of widely differing transmission quality. Other assessment methods, some of them more practical for day-to-day use and including, of course, some existing methods, can be fitted into the framework provided by the general method. These more specialized methods, once their position relative to the general background is clear, can be used with confidence and indeed can, within an appropriat narrower range of application, be developed specifically fo precision, repeatability or simplicity without leading to th danger of unrealistic results.

It must be emphasized in conclusion that assessment as dis cussed in the paper is primarily a laboratory method and doe not provide directly a good medium for discussion of value t the user. Consequently, although assessment results can some times be used directly at the research stage, great caution i needed if any attempt is made to forecast the merits of equipmen for specific applications. The main use of assessment method is indirect, namely as part of the process of rating speech links and it is the carefully combined use of rating and of user evalua tion[28] that can best provide adequate information on the trans mission aspects of speech links for operational use.

(6) ACKNOWLEDGMENTS

The authors are indebted to many present and former col leagues at the Post Office Research Station for their help especially in the course of prolonged experimental work. Ack nowledgment is made to the Engineer-in-Chief of the Post Offic for permission to make use of information contained in the paper

(7) BIBLIOGRAPHY

(1) FLETCHER, H.: 'The Nature of Speech and its Interpreta tion', *Journal of Franklin Institute*, 1922, **193**, p. 729.

(2) MARTIN, W. H.: 'Rating the Transmission Performance o Telephone Circuits', *Bell System Technical Journal*, 193 **10**, p. 116.

(3) MCKOWN, F. W., and EMLING, J. W.: 'A System of Effectiv Transmission Data for Rating Telephone Circuits', *ibid* 1933, **12**, p. 331.

(4) GRINSTED, W. H.: 'The Statistical Assessment of Standard of Telephone Transmission', *Engineering Supplement t the Siemens Magazine*, January, 1937, No. 140, p. 1.

(5) HUGHES, J. R.: 'Progress in the Study of Effective Trans mission by "Immediate Appreciation" Tests and Stati tical Assessment', *ibid.*, November, 1938, No. 162, p. 1.

(6) HUGHES, J. R.: 'The Comprehensive Assessment of Tele phone Communication Efficiency', *Journal I.E.E.*, 194 **89**, Part III, p. 195.

(7) POCOCK, L. C.: 'A Survey of the Telephone Transmissio Rating Problem', *ibid.*, 1948, **95**, Part III, p. 253.

(8) RADLEY, W. G., and SLACK, M.: Discussion on (7). *Ibid* pp. 266 and 269.

(9) SHANNON, C. E.: 'Mathematical Theory of Communication *Bell System Technical Journal*, 1948, **27**, pp. 379 and 62

(10) EGAN, J. P.: 'Articulation Testing Methods', *Laryngoscop* 1948, **58**, p. 955.

(11) SWAFFIELD, J., and DE WARDT, R. H.: 'A Reference Tel phone System for Articulation Tests', *Post Office Ele trical Engineers' Journal*, 1950, **43**, p. 1.

(12) LEE, B. S.: 'Some Effects of Sidetone Delay' (letter), *Journ of the Acoustical Society of America*, 1950, **22**, p. 639.

(13) RICHARDS, D. L.: 'An Outline of the Design and Analys of Experiments', *Post Office Electrical Engineers' Journa* 1951, **44**, p. 55.

(14) FINNEY, D. J.: 'Statistical Method in Biological Assa (C. Griffin and Co. Ltd., London, 1952).

(15) DE WARDT, R. H.: 'The Conduct of Articulation Measur ments', *Post Office Electrical Engineers' Journal*, 195 **44**, p. 159.

(16) RICHARDS, D. L., SHEARME, J. N., and BLUNDEN, R. M 'A Method of Generating Random Shapes for use Material in the Message Rate Method of Assessi Telephone Circuits', Post Office Research Repo No. 12970, 1952. (Restricted availability.)

7) RICHARDS, D. L., and BERRY, J.: 'Tables for the Generation of Logatoms based on Conversational Speech', Post Office Research Report No. 13596, 1952. (Restricted availability.)
8) RICHARDS, D. L., and SHEARME, J. N.: 'A Telephone Transmission Reference Circuit using Adjustable Interruption Distortion', *Post Office Electrical Engineers' Journal*, 1953, **45**, p. 175.
9) FLETCHER, H.: 'Speech and Hearing in Communication' (Van Nostrand, 1953).
20) RICHARDS, D. L.: 'Some Aspects of the Behaviour of Telephone Users as affected by the Physical Properties of the Circuit', p. 442 of 'Communication Theory', edited by Willis Jackson (Butterworths Scientific Publications, 1953).
21) RICHARDS, D. L., and SWAFFIELD, J.: 'Measurement of the Communication Rate of a Speech Link—a Preliminary Appraisal of the Problem', *ibid.*, p. 450.
22) 'Measurement of A.E.N. of a Commercial Telephone System (sending and receiving) by Comparison with the S.R.A.E.N.', C.C.I.F. Green Book, 1954, Vol. IV, 3.1.4, p. 68.
23) 'Immediate Appreciation Testing Method', C.C.I.F. Green Book, 1954, Annexes to Vol. IV, Annex 23.
24) BLYE, P. W., COOLIDGE, O. H., and HUNTLEY, H. R.: 'A Revised Telephone Transmission Rating Plan', *Bell System Technical Journal*, 1955, **34**, p. 453.
25) QUASTLER, H.: 'Studies of Human Channel Capacity', p. 361 of 'Information Theory', edited by Colin Cherry (Butterworths Scientific Publications, 1956).
26) RICHARDS, D. L., and ARCHBOLD, R. B.: 'A Development of the Collard Principle of Articulation Calculation', *Proceedings I.E.E.*, Paper No. 2143 R, September, 1956 (**103** B, p. 679).
27) Contribution of the Telephone Administration of the U.S.S.R. to Question 7 of the 4th Study Group, C.C.I.F., 1955–6, 4th S.G.—Documents No. 26 and 36.
28) SWAFFIELD, J., and RICHARDS, D. L.: 'Rating of Speech Links and Performance of Telephone Networks', (see page 65).

(8) APPENDIX

A number of terms are used in the paper, and elsewhere in ne field of assessment and rating, that merit an explanatory ote of the senses in which they are most valuable. Hard and ist definitions are not yet desirable or practicable in most of ne cases; still less is it desirable to exclude the terms from ther uses.

(8.1) Link, Path, Channel, Circuit and System

All five terms are used to cover the physical apparatus and onditions for speech communication between two people. It is iggested that the following practice should be adhered to herever possible.

Link or *speech link* is the most general term; it is applied ommonly but not exclusively to conditions of both-way transission. *Path* is the general term for use with one-way ansmission conditions; *channel* is used when the emphasis is n the physical equipment of a one-way path. Thus a microone connected to a telephone receiver provides a *channel*; if escriptions of such things as ambient noise, talking distance om the microphone and method of use of receiver are added, ne description of a *path* is completed. For both-way transission, *link* is the general term and *circuit* is used when the mphasis is either on the physical equipment alone (as compared ith the complete link) or on the two-way aspect of the ansmission.

(8.2) Reference Speech Link, Reference System

When a completely defined link is set up for measurement or reference purposes and includes a circuit and specified talking and listening conditions, it is known as a *reference speech link* or *reference system*; the latter term is in line with long established C.C.I.T.T. usage.

The *reference speech link* must also include a graduated variable impairment (e.g. an attenuator or signal/noise control) in the transmission path. A *high-quality reference speech link* is one in which high-quality components (especially transducers) are used so that there exists a setting of the variable impairment for which the transmission of the link is equal to that of the *ultimate reference speech link* (see below).

(8.3) Metre Air-Path

This is the rather idealized conception to which speech transmission standards can best be ultimately related. A *metre air-path* describes the transmission path provided when the talker's mouth and the listener's exposed ear, with monaural listening, are one metre apart in free space and in complete freedom from noise. The transmission is supposed one-way and the talker and listener have no means (e.g. visual) of communication other than that provided by the acoustic path.

(8.4) Ultimate Reference Speech Link or Ultimate Reference System

These terms are applied to that both-way link which is composed of two metre air-paths, as defined above, one in each direction. The conditions provided by this link are referred to as *ultimate reference conditions.*

(8.5) Listening Effort, Talking Effort

These are conveniently related to metre-air-path conditions; the first term implies effort relative to that which is needed by the listener when the talker uses, with one-way transmission, normal clear speech not degraded, as it would be in conversation, by the use of any knowledge the talker may acquire that the listener hears him without any difficulty owing to the perfection of the transmission conditions. In this sense the metre air-path provides the datum for listening effort; the talking effort under the conditions described is the datum for talking effort.

(8.6) Conversational Effort

The ultimate reference speech link provides the datum for *conversational effort* in the same sense that the metre air-path provides it for *listening effort*. Thus two normal subjects conversing over the ultimate reference speech link, or in practice in an ordinary quiet furnished room (which provides for this purpose inappreciably different conditions), require zero *conversational effort.*

(8.7) Standard Speech Link, Transmission Standard

Both these terms[28] refer primarily to working-type telephone apparatus representative of equipment in service. A *Standard Speech Link* includes specified ambient conditions at both ends and a variable impairment in the transmission path, one setting of which yields *datum transmission.* A *transmission standard* consists of the circuit (i.e. physical equipment) only and has no variable impairment.

It should be noted that, although there are some similarities in their use, *reference speech links* are not strictly 'Standard Speech Links', because they do not consist of practical working-type components and are not, in any case, in a form suitable for working trials and studies of user evaluation.

[The discussion on the above paper will be found overleaf.]

DISCUSSION ON THE ABOVE TWO PAPERS BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 27TH OCTOBER, 1958

Prof. E. C. Cherry: It was a classical view in electrical engineering that Fourier analysis and methods, which employ sine-wave characteristics, are complete and adequate. But this approach is clearly seen to be inadequate to describe the true *communicative* properties of channels; such tests measure only the properties of the apparatus itself and omit the most essential parts, namely the human beings at each end. Sine-wave response, or even transient characteristics, are only the start of a specification.

Perception involves us in a whole hierarchy of activities. At the lower end it is true that sine-wave characteristics and signal analysis form one category of specification. But if my telephone has adequate frequency characteristics, not only is it able to speak to me in good English, it can also make a host of other sounds which I shall never want. Therefore, we have to step up to the next category and make tests with those particular classes of sound which I *do* want, namely the phonetic category. This involves not only physical signals but essentially the *relationship* between these and the listener. Going further, I want my telephone to talk to me, particularly, in English phonetics. Therefore, we want tests of another category, which bring in the particular habits of the speakers and users. Again, if my telephone is to speak sense at me, I must devise my tests so that they involve the speaker's predictions of what my response will be and my prediction of what he wants me to think and do. Therefore, it must be emphasized that it is not possible to specify the properties of a telcommunication system by a single class of measurement; a whole range is required.

I agree with the authors that, in the top category, one wants measures of behaviour, because conversation is a mode of behaviour; the execution of the puzzles which the authors describe represents one good example, but there are many others.

In the last two years I have been engaged on a crusade for The Institution, going round to almost every Centre in the British Isles, trying to arouse interest in the need for widening engineers' concepts, to include especially some of those of the behavioural psychologist, because some of their methods and techniques are very relevant to our field, but are not as widely known among engineers as they ought to be. Engineers regularly aim to stretch their familiar concepts beyond the legitimate sphere, but there are other concepts which can profitably be included and which also lead to objective measurements.

In this connection, Fig. 4A of Paper No. 2605 involves the questionable concept of thought setting up words. The majority of our speech has little or nothing to do with thought, but is, in fact, response to motor habits. The idea of thoughts 'setting up' speech, like a coding process, is mistaken; neither should we look upon perception of speech as a kind of decoding process. We cannot talk about 'channel capacity' in the present context, because no coding is involved. The capacity theorem involves the concept of maximum information rate over all conceivable forms of encoding. But here speech is taken as raw speech, and no coding is involved.

We have a long way to go yet in this whole field of study, by stretching our engineering concepts very carefully and, while realizing the limits of this stretching, bringing in radically new concepts when required, for which psychologists are very ready to help with objective methods. The difficulty in this field of work is not, as many people assume, that subjective matters are involved, for all observation is subjective.

Dr. G. W. Sutton: With regard to Paper No. 2666, 30 years ago it was usual to set up working standards such as the one described. It later became the practice to set up a 'perfect' system, incorporating a condenser microphone and moving-co receiver, to which various forms of distortion were added. Th return to the simpler form of standard will surely be welcome by workers in this field.

In Paper No. 2605 it is stated that testing methods such a immediate appreciation are 'ineffective collectively or individuall in giving any significant results'. I do not think that the author will be surprised if I disagree with this very sweeping statement It can be claimed that immediate appreciation, used in the simpl form originally suggested by Mr. Grinsted, gives most useful an significant results. Like other such methods it can be, and ha been, elaborated and refined until it loses its original appeal an value.

In Section 4, it is stated that 'the essence of a sentence is it meaning rather than its separate words, and this idea is exploite in the immediate-appreciation method'. The authors clearl have not appreciated the idea underlying the method, as explaine in References 5 and 6 of the paper. Complete sentences are not in fact, used, and the testing crew are asked to mark as a 'success only those phrases that are immediately recognized, withou mental effort of any sort. It is the words rather than thei meaning that they are asked to recognize.

Mr. W. West: The subjects used for obtaining the mean opinion scores are civil servants; their opinions deserve respect but they may not be representative of telephone users as a whole I regard the scores not as having any absolute value, but as useful for comparisons only. In this respect they do not differ from other subjective assessment scales. The 'over 60's' comprise a large group of users that is practically unrepresented in the tests. Their hearing is generally less keen, and their opinions might throw more emphasis on signal strength than on signal/noise ratio as the important factor (e.g. Section 3.2).

In Section 5.3 it is stated that 45 dB may be taken as a mean value of room noise. This may be reasonable for expensive calls, for which some precautions might be taken to make or take the call in quiet surroundings, but it is a very low level for the many telephones within earshot of our heavy and widespread road traffic. In any case, the mean value of room noise is not the important quantity. If we take the case, for example, of noise due only to someone else talking in the room, the mean level at the telephone may be less than 50 dB while the louder vowel sounds rise above 60 dB, and these occur so frequently as to be almost as impairing as if they were continuous.

The formal definition of the transmission performance of a circuit (i.e. 'the effectiveness of the circuit for transmitting or reproducing speech in the circumstances in which it is used') shows the arbitrary nature of any single rating of a circuit which may be used in many different circumstances. Hitherto the aim of planning, so that conversation should be practicable between any two stations that can be connected, has been concentrated on assessments made in conditions taken to represent the trans mission limit. These ratings are used to avoid changes in the network which would depart from the aim, and they have the power of veto. The papers now introduce the concept of assessing the transmission performance of a whole network Have any exercises been carried out in the application of this concept. For example, how would Table 2 be changed if it were assumed that all the telephone sets used were of the newest type?

Mr. G. W. Barnes: The authors are mainly concerned with the telephone network, but other applications are mentioned. Can the same arguments be applied to radiotelephones and, in par ticular, to air-ground radiotelephones to which reference is made

the papers? The techniques described appear to assume irly steady-state conditions, which do not always exist on air-ound radiotelephone circuits. On v.h.f. transmissions the gnal strength varies with changes in range and aircraft attitude, id on h.f. links, fading and distortion occurs as a result of ulti-path propagation. Do the authors consider that these fects can be balanced against another form of degradation in standard circuit? At the Royal Aircraft Establishment we ave compared radiotelephone systems by measuring sentence, ord and figure intelligibilities from recordings made during ght. It would appear that this method is satisfactory for com-arison purposes, but must not be used to assign a figure of merit a particular system. On the other hand, it would be difficult apply some of the assessment methods based on the mental fort required for a particular task.

The improvement which results from an exchange of assess-ents on a telephone circuit is mentioned. This is also true on diotelephones in that operators normally start with such an change, the reports being given on a readability scale 1 to 5 milar to the assessment scale given in the paper. Inevitably e man who is least understood tries harder.

I should like to have examples of X and Y factors to indicate e range of effects which have been considered. One particular fect is that of jamming by other signals. This occurs on tele-none circuits owing to other people speaking in the same room r to crosstalk, and on radio circuits it may be due to inadequate ceiver selectivity, spurious responses or non-linearity. Can the sceptibility of equipments to jamming be taken into account hen making assessments?

Mr. T. H. Flowers: The papers come at an opportune time in e development of electronic exchanges. Existing exchanges in transmission lines together with metallic contacts which have effect on transmission. Electronic exchanges make such con-ections with apparatus which has properties similar to those of transmission line. The attenuation introduced may have, epending on the system, a high mean value (e.g. 3 dB) and small ndom variation, or zero mean value and high random varia-on, e.g. standard deviation 1·5 dB. The papers show that ndom variation of line attenuation about a mean is of very ttle importance to speech transmission compared with variation the mean. This is best seen from Fig. 7 of Paper No. 2065. he curves represent a standard deviation of about 12 dB to hich the r.m.s. addition of even 3·5 dB due to line variations ould add only 0·5 dB to the total. Hence the curves are odified by passing through the same points on the 50th per-entile, but at slightly decreased slope. A variation in mean loss ifts the rating scale by an equal amount, and this has a greater fect.

An early design of electronic exchange system provided for cal calls to be degraded by about 6 dB and the longest calls to e improved by about 12 dB. This was thought to give an verall improvement in service, but Table 2 of Paper No. 2666 ows that it will not.

I feel doubtful about the validity of the practice of mean-inion scores. The fact that it gives results which are plausible es not necessarily mean that they are significant. The major ntribution to any particular mean-opinion score must come om one, or at the most, two levels of opinion. The scores are erefore not much more stable than the primary data, and often ss useful. For example, I believe that my first point can be educed only from the primary data.

Mr. R. Wakar: No reference is made in the papers to the ssessment of transmitted-speech naturalness. This feature gains rominence when the introduction of coded speech-transmission ystems is considered. It is likely to have a considerable share the determination of users' satisfaction.

Halsey and Swaffield* refer to the 'inhuman' quality of synthesized speech and to the necessity for the quantitative measurement of naturalness. According to that paper, synthesized speech, apart from being very unnatural, can otherwise provide a high articulation efficiency.

Mr. A. C. Timmis: To discover a satisfactory criterion of telephone efficiency is a problem which has baffled the C.C.I. for many years, and therefore the old limit of 40 dB reference equivalent still stands. This is purely a volume (loudness) standard and was based on the judgment of a few experts, talking over an artificial line with the c.b. telephones of 30 years ago. The quality was atrocious compared with modern telephones and side tone could be deafening.

If interpreted literally, therefore, the 40 dB standard relative to S.F.E.R.T. could be very poor, but fortunately modern working standard telephones, calibrated for volume against S.F.E.R.T., have greatly superior quality. Therefore we actually work to a much better overall standard, namely S.F.E.R.T. volume plus modern quality. Obviously this standard should be maintained for the present. Paper No. 2666 briefly mentions the Post Office standard subscriber's line (450 ohms of 10 lb cable) and more realistic methods than volume comparison for rating subscribers' lines against the standard.

In 1946 the Post Office adopted this standard both for volume (reference equivalent) calculations and for effective transmission rating. At that time we were seeking a working criterion of performance by various subjective methods, none of which was fully satisfactory. Nevertheless it was found possible to rate various subscribers' lines against the standard with reasonable accuracy. Undoubtedly these results are nearer the truth than the old volume method, which assumes that volume depends on d.c. resistance, disregarding the effects of side tone. Calculated on this basis the design limit for 20 lb cable is too long, and for $6\frac{1}{2}$ lb cable it is too short. By definition, both methods are in agreement for 10 lb cable. Thus the authors' 'more realistic' method of designing subscribers' lines ensures that the present high standard of transmission is maintained.

When the C.C.I. recommendations are applied in this way to planning the telephone network of an under-developed country, where the existing plant may be disregarded, the maximum reference equivalent between subscribers (with type 700 telephones) is about 27 dB. This is practically the same as in the Post Office network, which has been proved by experience to give a very satisfactory general standard of transmission. Hence it appears safe to plan on this basis while the perfect criterion is being sought.

Do the authors consider that a purely subjective, rather than an objective, method is likely to provide the answer, in view of the possible use of vocoders and kindred devices in the trunk network?

Mr. C. S. Mills (*communicated*): For two years, we, in the Ministry of Supply, have been engaged with the authors on the establishment of a Transmission Standard for use in planning Army communication systems. For such collaboration between different laboratories to be effective there must be close agreement on such things as the objective measurements of transducers, design and specification of reference systems and methods of use, and also on terminology. I was therefore particularly pleased to see the list of terms given in the Appendix to Paper No. 2605, although I cannot agree that 'hard and fast definitions are not yet desirable'. For example, the terms 'Standard Speech Link' and 'Reference Speech Link' should be kept quite distinct.

During an assessment the percentage of sentences understood without effort (immediate appreciation) is higher after conversa-

* Halsey, R. J., and Swaffield, J.: 'Analysis-Synthesis Telephony, with special reference to the Vocoder,' *Journal I.E.E.*, 1948, 95, Part III, p. 391.

tion than before [Paper No. 2605, Fig. 5, curves (*s*) and (*q*)]. Is there any clear evidence of 'learning' taking place as the first immediate-appreciation test proceeds, or is the improved score in the second immediate-appreciation test entirely attributable to the fact that a conversation between the two subjects has taken place in the intervening period?

The second question concerns the dilemma posed by the economic importance attaching to the correct selection of the Transmission Standard, on the one hand, and the admitted lack of precision with which this selection can be done, on the other. Can the authors say at what level of responsibility in the Post Office decisions are made relating to changes in the Standard?

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Mr. D. L. Richards and **Dr. J. Swaffield** (*in reply*): We agree substantially with Prof. Cherry's remarks and hope that the following comments, in conjunction with a study of Section 2 of Paper No. 2605, will remove his doubts about Fig. 4A. It is true that the idea-bearing element is a minute part of all the physical content of speech, but nevertheless it is the existence of this part at the two extreme ends, A and A′ in Fig. 4A, which provides the essence of meaningful conversation. Prof. Cherry's remarks, if applied to conversation, suggest motor activity of the talker influenced neither by that of the other talker nor by his own new 'thought'; this type of 'conversation' would not really need a telephone line at all. Incidentally only an excessively narrow definition of coding would exclude its application to the process taking place between A and B, but this has, in any case, no relevance to our use of the term 'channel capacity', which is quite orthodox and conforms with Prof. Cherry's remarks. The term is applied to the physical channels (as shown in the Figures), and these do not adjoin the relatively controversial world to the left of C or to the right of C′.

We agree with Dr. Sutton that the immediate-appreciation principle, as originally formulated by Grinsted and Hughes* (i.e. that the listener is required to 'record the occasions when any sort of mental effort or deduction is required for the appreciation of the meaning of the sentence heard') is very useful over its somewhat limited range of application. There is no conflict between the use of Reference Systems and Standards; both are still needed as they have been for the last 30 years. As Mr. Mills emphasizes, confusion between the two often occurs.

Mr. West's remark that, for persons having high thresholds of audibility, the range over which $\alpha + \beta = 1$ is restricted, illustrates a difficulty which arises when X and Z factors interact, and emphasizes the importance of a suitable choice of Standard and assessment method. We agree that new equipment must be rated relative to the old (the Standard) over the whole relevant range of ambient conditions because ratings vary considerably over this range.

Several studies of network transmission performance have already been made (one is referred to by Mr. Flowers), and broadly the effect of varying the sensitivity is to increase the ratio

$$\frac{\%\text{ unsatisfied subjects}}{\%\text{ satisfied subjects}}$$

by the factor 2 for every 3dB decrease in sensitivity.

In reply to Mr. Barnes, the principles covered by the papers apply to speech links generally, including radiotelephones, but we have no direct experience of air-ground links. Unsteady conditions cause difficulties of application rather than principle, and an example involving slow fading is quoted at the end of Section 5.4 of Paper No. 2666. We feel that the tests at R.A.E. tend to confuse user evaluations and assessments for rating; this happens readily when the latter (laboratory tests in essence) have to be conducted partly in an operational setting because conditions cannot be sufficiently simulated in the laboratory. It would be interesting to go through the procedure discussed in Section 5.1 (some adaptation to the air-ground problem would be needed); this should expose any essential difficulties. The difference between X and Y factors is one of convenience; those under the control of the designer or simply related to design are handled differently from those that are not. Circuit noise is usually an example of the former, but ambient noise is one of the latter.

We do not agree with Mr. Flowers that experiments can be analysed efficiently in percentage form without calculating mean-opinion scores. The latter do indeed contain more information than any single percentage; for example, at a mean score of 2·5 the five percentages are approximately 8, 42, 42, 7 and 0·4%, so that no fewer than four categories make appreciable contributions.

In reply to Mr. Wakar, the value of specialized performance descriptions is during research into new types of speech transmission (an example is the use of a 'naturalness' measure in synthetic speech research). This is covered in Section 1(*d*) of Paper No. 2605, but as is made clear in Section 1(*e*) and elsewhere, such measures have little direct value at the later stages when complete equipments are being rated and their use planned.

Capt. Timmis points out one reason for Transmission Standards tending on revision to improve slightly; another reason is that, whenever a new telephone set is introduced, the one it supersedes is usually the commonest in the network and so conveniently becomes the revised Standard. The need to retain satisfactory performance on calls between new and old sets then tends to result in a slight improvement in Standard. His queries about subjective and objective methods are fully covered in Section 3.4 of Paper No. 2666 and Section 4 of Paper No. 2605.

In reply to Mr. Mills, there is evidence of a slight 'learning' effect during a listening immediate-appreciation test, but this is entirely insufficient to account for the increase after conversation. Listening opinions, moreover, show no systematic increase in score as the test proceeds. A Transmission Standard cannot be set up in final form by any single administrative decision; it is evolved by studying what has proved workable under operational conditions either in the network in question or in a roughly similar one. If some entirely new operational use is being considered and no even remotely similar network exists, then, as a last resort, a condensed and specialized user trial can be made. Careful thought will show that the procedures discussed in Sections 5.1 and 6 of Paper No. 2666 ensure that Standards once initiated tend to be self-correcting.

* See Reference 5 of Paper No. 2605.

.396.965.4:621.396.677.852

The Institution of Electrical Engineers
Paper No. 2735 R
Nov. 1958
©

HE USE OF DISPERSIVE ARTIFICIAL DIELECTRICS IN A BEAM-SCANNING PRISM

By J. S. SEELEY, B.Sc.(Eng.), Ph.D., Graduate, and J. BROWN, M.A., Ph.D., Associate Member.

(*The paper was first received* 19*th February, and in revised form* 15*th July*, 1958. *It was published in November*, 1958, *and was read before the* RADIO AND TELECOMMUNICATION SECTION 24*th November*, 1958.)

SUMMARY

The direction of a radiated beam can be altered by passing it through prism. If the prism material is dispersive, the final direction is quency sensitive and beam scanning can therefore be achieved by quency modulation. Two types of dispersive artificial dielectric e considered for use in such a prism, one consisting of an array of ds, and the other of an array of sheets containing a pattern of onant slots. Measured values of the electrical constants of both ays over a wide range of wavelengths are included, and the properties the beam-scanning prism are described in detail.

LIST OF PRINCIPAL SYMBOLS

a = Spacing between elements of artificial-dielectric array in direction parallel to H-vector.
b = Spacing between elements of artificial-dielectric array in direction of propagation.
c = Spacing between elements of artificial-dielectric array in direction parallel to E-vector.
f_r = Resonant frequency of sheet of slots.
i = Angle of incidence.
n = Refractive index of isotropic material.
n_0 = Refractive index of anisotropic material at normal incidence.
n_i = Refractive index of anisotropic material at angle of incidence i.
Q_L = Loaded Q-factor of sheet of resonant slots.
$Z = 1/Y$ = Wave impedance of supporting dielectric material.
Z_0 = Free-space wave impedance.
ϵ_r = Relative permittivity of supporting dielectric material.
λ = Wavelength in supporting dielectric material.
λ_0 = Free-space wavelength.
λ_r = Resonant wavelength of sheet of slots within supporting dielectric material.
$, \lambda_2$ = Lower and upper cut-off wavelengths within supporting dielectric material for propagation in the first pass band.

(1) INTRODUCTION

In many radar applications, the direction of the radiated am has to be scanned over a wide range of angles. The anning can always be done mechanically by rotating the whole rial system, but this places a limit on the rapidity of the scan, rticularly when a narrow beam requiring a large aerial is volved. Many forms of scanner have been developed in which ly a part of the aerial need be moved to change the beam rection, and these represent a great advance on moving the mplete aerial. For very fast scans, the ideal solution is an rial system in which no moving parts are needed to change e beam direction. Such a system can be achieved in principle by a dispersive prism, as shown in Fig. 1. The wave entering the prism is a beam collimated by any suitable reflector or lens aerial, and is refracted by the prism according to the laws of geometrical optics. The direction of the emerging beam depends on the refractive index of the prism material, n, through the equation

$$n = \frac{\sin (P + D)}{\sin P} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

P being the prism angle, and D being the deviation in the beam direction produced by the prism. For dispersive materials n depends on frequency, and with the arrangement of Fig. 1 the

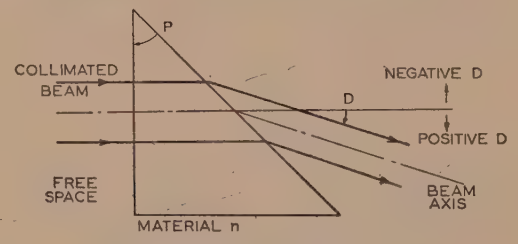


Fig. 1.—Refraction through a prism.

direction of the radiated beam can be altered by changing the frequency of the input signal, which gives a scanning system with no moving parts. An essential requirement is a dispersive material from which to make the prism. No conventional dielectric has a sufficient degree of dispersion, and the work described in the paper was carried out as an attempt to determine the behaviour of artificial dielectrics[1] in this application.

Previous studies of artificial dielectrics have shown, in general terms, the factors determining dispersion, and attention has been drawn to two possible mechanisms[2] by which dispersion is caused. In the first, known as 'cavity dispersion', the metal elements forming the artificial dielectric are spaced apart in the direction of wave propagation by a distance of the order a half wavelength, so that the dielectric is, in effect, a series of direct-coupled resonators. The behaviour with respect to frequency is very similar to that of the direct-coupled resonator type of microwave filter. The second possible mechanism of dispersive behaviour, called 'dipolar dispersion', is caused by a resonance in the reflecting properties of the elements forming the array and is analogous to that in conventional dielectrics. The two causes of dispersion are not exclusive, but, in general, it is possible to regard one of them as the dominant factor. As is well known from the theory of dielectrics, dispersion is always associated with relatively high levels of attenuation. The design of a material suitable for the prism in Fig. 1 calls not only for high dispersion but also for low attenuation, and the objective is therefore the largest possible dispersion for a given tolerable attenuation. This considerably complicates the work, for although it is relatively easy to predict the dispersive properties of artificial dielectrics with sufficient accuracy, the attenuation depends on so many factors

Dr. Seeley was formerly in the Electrical Engineering Department, Imperial College Science and Technology, University of London, and is now at the Royal Aircraft ablishment.
Dr. Brown is in the Electrical Engineering Department, University College, London.
The paper forms part of a thesis approved for the degree of Doctor of Philosophy London University.

that theoretical prediction is inaccurate and measurement is virtually essential.

In the paper two possible dispersive artificial dielectrics are investigated in detail, the first consisting of an array of parallel rods and the second an array of conducting sheets in which are cut resonant slots. These correspond essentially to the 'cavity' and 'dipolar' types, respectively.

(2) RODDED DIELECTRICS

(2.1) Theory

The rodded dielectric is an example of a phase-advance array[3] for which the refractive index is less than unity. Consideration of the group velocity shows that such arrays are necessarily dispersive. The theory for rodded dielectrics, which consist of arrays of rods whose axes are parallel to the direction of the incident electric field, is developed from an equivalent trans-

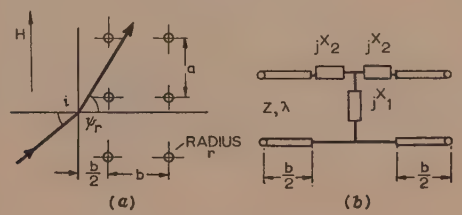


Fig. 2.—Rodded dielectric.
(*a*) Diagram.
(*b*) Equivalent circuit.

mission line circuit [Fig. 2(*b*)]. The results required for the present investigation will be summarized in this Section. All attenuation effects are neglected in the analysis.

A plane wave with its electric field parallel to the rods is assumed to be incident from free space on to the rodded dielectric, the angle of incidence being i [Fig. 2(*a*)]. The rods are embedded in a solid dielectric of relative permittivity ϵ_r which extends a distance $b/2$ from the mid-plane of the first grid of rods. The effective refractive index of the rodded dielectric is given by

$$n_i^2 = n_z^2 + \sin^2 i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where n_z is given by

$$\cos\left(\frac{2\pi b}{\lambda_0}n_z\right) = \left(1 + \frac{X_2}{X_1}\right)\cos 2\phi + \frac{[Z^2 - \cos^2\psi_d(X_2^2 + X_1X_2)]}{2X_1Z\cos\psi_d}\sin 2\phi \quad . \quad (3)$$

in which $\phi = \frac{\pi b}{\lambda_0}\sqrt{(\epsilon_r - \sin^2 i)}$ and $\cos\psi_d = \sqrt{\left(1 - \frac{\sin^2 i}{\epsilon_r}\right)}$. X_1 and X_2 are the reactances in the equivalent circuit of Fig. 2(*b*) and are tabulated in Reference 4. X_1 is inductive and X_2 is capacitive. For thin rods, i.e. $r \ll a$, X_2 is negligible and X_1 is given sufficiently accurately by $\sqrt{\epsilon_r}Z\frac{a}{\lambda_0}\log_\varepsilon\left(\frac{a}{2\pi r}\right)$, so that eqn. (3) reduces to

$$\cos\left(\frac{2\pi b}{\lambda_0}n_z\right) = \cos 2\phi + \frac{\lambda_0 \sin 2\phi}{\sqrt{\epsilon_r}2a\cos\psi_d\log_\varepsilon\left(\frac{a}{2\pi r}\right)} \quad . \quad (4)$$

The wave propagating in the rodded material travels in a direction making an angle ψ_r with the z-axis, where

$$\tan\psi_r = \frac{\sin i}{n_z} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The above equations completely specify the refracting properties of the material.

A further property of great importance in practical applications is the reflection coefficient, ρ_i, at the interface between free space and the material, and this is given by

$$\rho_i = \frac{Z_{ni} - Z_0\sec i}{Z_{ni} + Z_0\sec i} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where Z_{ni}, the effective wave impedance of the material for incidence at angle i, can be calculated from

$$Z_{ni} = \frac{Z_0}{\sqrt{(\epsilon_r - \sin^2 i)}} \times \frac{\left(\sin\phi + \frac{X_2}{Z}\frac{\cos\phi}{\cos\psi_d}\right)\left[2\cos\phi + \frac{1}{X_1}(Z\sin\phi\cos\psi_d + X_2\cos\phi)\right]}{\sin\left(\frac{2\pi b}{\lambda_0}n_z\right)} \quad . \quad . \quad . \quad . \quad (7)$$

where ϕ and $\cos\psi_d$ are the same quantities as those used in eqn. (3). Eqn. (7) simplifies in the case of thin rods to

$$Z_{ni} = \frac{Z_0}{\sqrt{(\epsilon_r - \sin^2 i)}}\frac{\tan\phi}{\tan\left(\frac{\pi b}{\lambda_0}n_z\right)} \quad . \quad . \quad . \quad (8)$$

The above relations apply for any angle of incidence, including normal incidence when $i = 0$. The refractive index, wave impedance and reflection coefficient at normal incidence will be denoted by n_0, Z_{n0} and ρ_0, respectively.

The quantity of most interest for the present investigation is the dispersion or rate of change of refractive index with operating wavelength. The dispersion can be examined in curves of refractive index plotted against wavelength, the simplest case to consider first being that of thin rods with a normally incident wave. The refractive index equation is then rewritten as

$$\cos\left(\frac{2\pi b}{\lambda_0}n_0\right) = \cos\left(\frac{2\pi b}{\lambda_0}\sqrt{\epsilon_r}\right) + \frac{\lambda_0}{\sqrt{\epsilon_r}\lambda_c}\sin\left(\frac{2\pi b}{\lambda_0}\sqrt{\epsilon_r}\right)\tan\frac{\pi b}{\lambda_c} \quad . \quad . \quad . \quad . \quad (9)$$

where λ_c is the longest wavelength in the material for which propagation can occur and is given by the smallest root of the equation

$$\tan\frac{\pi b}{\lambda_c} = \frac{Z}{2X_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

or

$$\frac{2\pi b}{\lambda_c}\tan\frac{\pi b}{\lambda_c} = \frac{\pi b}{a}\frac{1}{\log_\varepsilon\frac{a}{2\pi r}} \quad . \quad . \quad . \quad . \quad (11)$$

Propagation can occur for all free-space wavelengths between $\sqrt{\epsilon_r}2b$ and $\sqrt{\epsilon_r}\lambda_c$, the refractive index being $\sqrt{\epsilon_r}$ and zero at these two wavelengths, respectively. In Fig. 3, the refractive index is plotted against $\lambda_0/\sqrt{\epsilon_r}\lambda_c$ for various values of the parameter b/λ_c.

The general shape of the refractive-index curves is not greatly different when the eqn. (3) appropriate to thick rods is used. The essential difference is found by determining the cut-off wavelengths within the material, which are given by the roots of the following equations:

$$\tan\frac{\pi b}{\lambda_2} \simeq \frac{Z}{2X_1 + X_2} - \frac{X_2}{Z} \quad . \quad . \quad . \quad . \quad (12)$$

and

$$\lambda_1 \simeq \frac{2b}{\left(1 - \frac{2X_2}{\pi Z}\right)} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

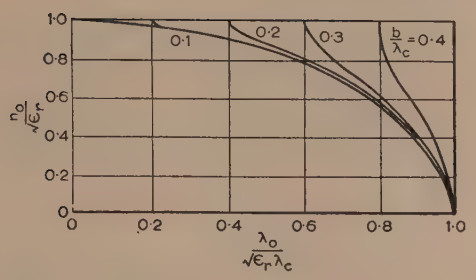


Fig. 3.—Refractive index of rodded dielectric as a function of wavelength.

om the first of these, computation shows that the free-space t-off wavelength for $n_0 = 0$ is nearly equal to $\sqrt{\epsilon_r}\lambda_c$ as cal-lated from eqn. (10). The second shows that the smaller t-off wavelength is reduced from the value for the thin-rod se, $\sqrt{\epsilon_r}2b$. This implies that the use of thick rods extends the ss band from that obtained with thin rods having the same acing b and the same value of X_1. In addition, the refractive dex at the lower cut-off wavelength is less than $\sqrt{\epsilon_r}$. The spersion must therefore be reduced when thick rods are used, d thus it is desirable to keep the rods sufficiently thin for the mpler theory to apply.

(2.2) Measurement of the Dispersive Properties

It is clear from Fig. 3 that the largest dispersion is obtained using a value for b/λ_c as near to 0·5 as possible. The value $\lambda_c = 0{\cdot}4$ was selected for a practical investigation of this and e other important properties—surface reflection and attenua-on. The spacing of the array was made symmetrical (i.e. $a = b$), that a 45° prism could be designed with the rods in successive ids aligned opposite to each other (see Fig. 4). Perspex was

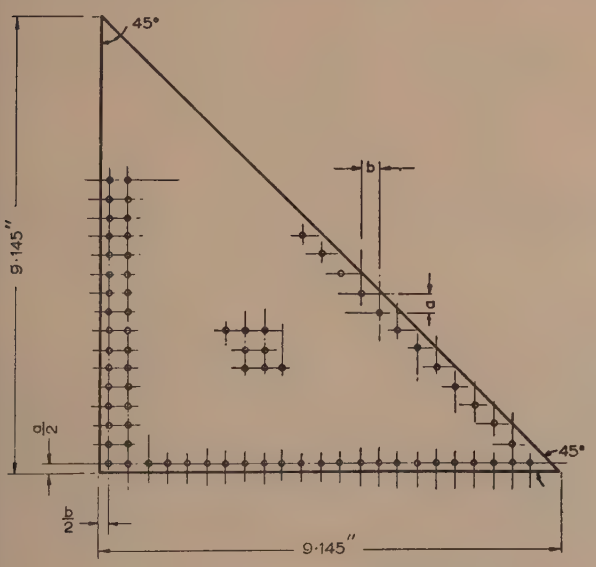


Fig. 4.—Dimensions of rodded prism.

Not to scale.
$a = b = 0{\cdot}188$ in.
Holes drilled No. 70 (0·028 in. dia.), clearance for No. 23 s.w.g. (0·24 in).
terial, Perspex, 0·192 in thick.

ed for the base material, and the wavelength at which $= 0{\cdot}5\sqrt{\epsilon_r}$ was chosen as the mid-band wavelength. The mensions of the prism are shown in Fig. 4; the array of over 3 000 holes involved was jig-bored to ensure an accuracy of $\pm 0{\cdot}001$ in in the positioning of each hole. Rods were cut with a special shearing tool to a length such that they projected 0·002 in from each surface of the prism, this slight projection ensuring a good biting contact with a tin-foil coating. A strip of length 22 cm and width 2·4 cm containing seven rows of holes was also machined in polystyrene.

Measurements were made in a 1 cm parallel-plate spectrometer.[5] The prism was used at normal incidence to obtain values of the real refractive index[5] and the modulus of the reflection coefficient.[6] Transmission and reflection characteristics of the type shown in Fig. 5 were obtained from measurements at

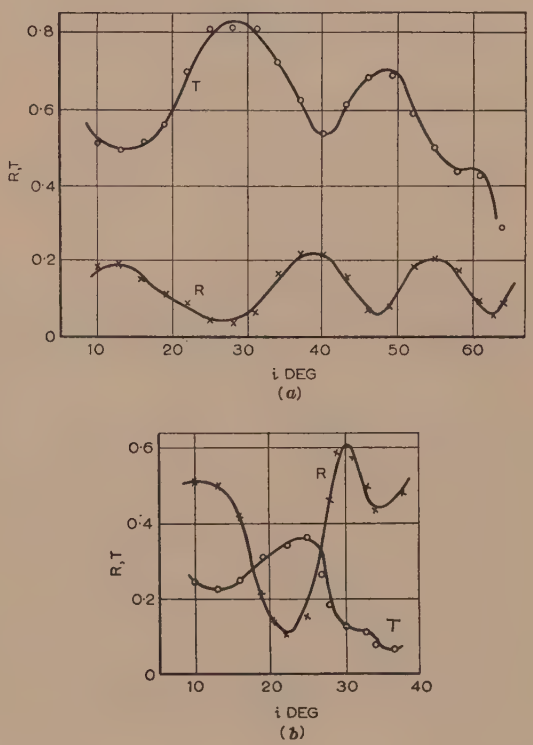


Fig. 5.—Power reflection and transmission coefficients, R and T, of rodded dielectric strip at varying incidence.

(*a*) $\lambda_0 = 1{\cdot}035$ cm.
(*b*) $\lambda_0 = 1{\cdot}16$ cm.

varying incidence with the strip. The maximum and minimum values of these characteristics, and the angular positions at which they occur, were used to obtain values of the complex refractive index and complex reflection coefficient at various angles of incidence.[7] The refractive indices obtained in this way showed a slight degree of anisotropy, an effect which will be discussed in the next Section. Experimental and theoretical values of the real refractive index, the attenuation and the modulus of the reflection coefficient are shown in Fig. 6. The phase angle of the reflection coefficient, as measured, was in all cases less than 5°.

The measured values of n_0 are in excellent agreement with the theoretical values in mid-band, but the dispersion at the edges of the pass-band is seen to be reduced even further than predicted by eqn. (3).

The attenuation characteristic (d) in Fig. 6 has been calculated from an equation recently developed.[8] For thin rods of commercial copper, at a frequency of 27 Gc/s, this may be expressed as

$$\text{Attenuation} \simeq 0{\cdot}001\sqrt{\epsilon_r}\frac{a}{2\pi r}\left(\frac{Z}{X_1}\right)^2 \times \frac{\sin\left(\frac{2\pi b}{\lambda_0}\sqrt{\epsilon_r}\right)}{2b\sin\left(\frac{2\pi b}{\lambda_0}n_0\right)} \text{ dB/unit length} \quad . \quad (14)$$

The measured attenuations are approximately seven times larger than the theoretical values, and can be associated solely with the conducting array, since the loss in the supporting material,

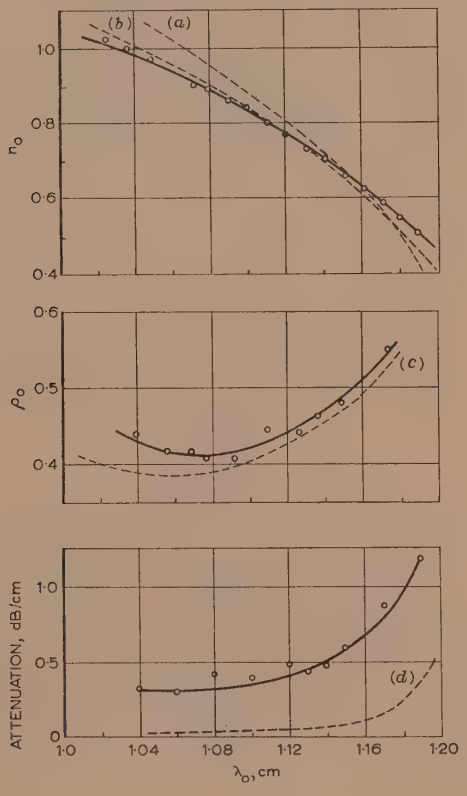


Fig. 6.—Theoretical and experimental characteristics of rodded dielectric.

—°———°— Measurement.

(*a*) Calculated from simple theory [eqn. (9)].
(*b*) Calculated from exact theory [eqn. (3)].
(*c*) Calculated from eqn. (7).
(*d*) Calculated from eqn. (14).

polystyrene, can be ignored. The following are possible reasons for this discrepancy:

(*a*) Eqn. (14) applies only to arrays with thin rods.
(*b*) The r.f. surface conductivity of the rods may be reduced by contamination from the solution of Vaseline in carbon tetrachloride used to attach the tin-foil to the top and bottom surfaces of the prism.
(*c*) The presence of only small amounts of r.f. contact resistance at the ends of the rods could absorb the additional power lost.

To check the last point the d.c. contact resistance of the rods has been measured by means of a Kelvin double bridge. Pressure pads 1·8 cm in diameter were used to isolate circular sections of the rodded material, with or without tin-foil, containing 32 rods. The d.c. contact resistance of each rod when covered with tin-foil at each end was found to be 0·002 5 ohm, which is approximately seven times the d.c. resistance of the rod. The resistance of each rod at 27 Gc/s is 0·11 ohm, and the total resistance required to produce the measured attenuations is approximately 0·8 ohm. The additional resistance could easily be due to a combination of r.f. contact resistance at the ends of the rods and a reduction in the surface conductivity

The series reactance X_2 must be very much less than the shunt reactance X_1 if an array with a large value of b/λ_c is not to suffer from a reduction in dispersion for the reason discussed at the end of Section 2.1. This is accomplished by reducing r and a while increasing the ratio a/r, to the smallest permissible value (given in Reference 4) for which $X_2 \ll X_1$. Eqn. (14) shows that these reductions in the dimensions of each grating will increase the attenuation, since the value of $a/2\pi r$ has been increased while the original values of X_1, b/λ_0 and n_0 have been retained. This suggests that r should be made equal to the largest value for which the simple theory applies.

(2.3) Measurement of the Properties at Oblique Incidence

The values of n_i and ρ_i, measured at 1·10 cm, are shown in Fig. 7. The behaviour of the array is more isotropic than

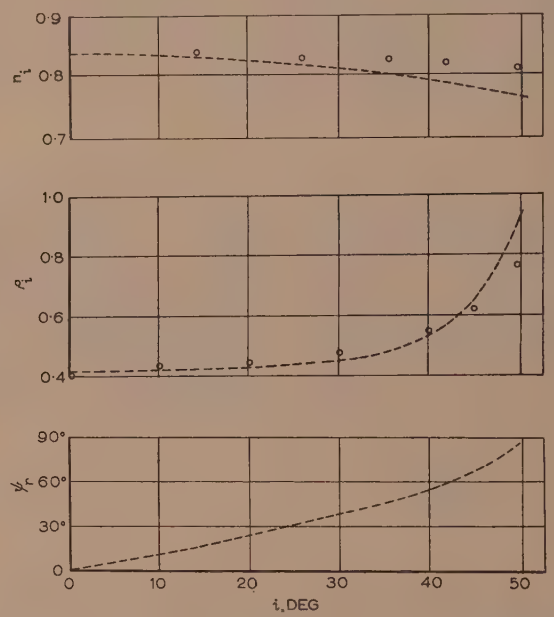


Fig. 7.—Characteristics of the rodded dielectric at oblique incidence.

$\lambda_0 = 1{\cdot}10$ cm.
– – – – Calculated values.
O Measured values.

expected in practice (i.e. the properties do not vary as much as expected with variations in the direction of propagation), and this feature is especially marked at angles of incidence greater than 45°. The refractive indices at $i = 90°$ and $i = 0$ must be equal because of the symmetry of the array, so that departures from the simple behaviour are expected at and above $i = 45°$.

The effect of the attenuation of the material on the value of n_i near the critical angle of incidence has been examined with the aid of the following approximate expression,[7] which applies to isotropic materials:

$$n_i \simeq n_0\left\{1 + \frac{k_0^2(k_0^2 + \sin^2 i)}{2[n_0^2(1 - k_0^2) - \sin^2 i]}\right\} \quad . \quad . \quad . \quad (15$$

where k_0 is the index of absorption at $i = 0$. Computation from this equation with the measured value of k_0 shows that $n_i = 1{\cdot}005\, n_0$ at $i = 49°$ (2° removed from the critical angle). It may therefore be concluded that a small part of the discrepancy between the theoretical and measured values of n_i at large

gles is due to the attenuation of the array. It is of interest it, because of its low refractive index, the rodded dielectric one of the few materials in which the behaviour indicated by n. (15) will be detectable. Eqn. (3) should be reformulated king allowance for attenuation before the behaviour of the dded dielectric at oblique incidence can be fully understood.

(3) RESONANT SLOT DIELECTRICS

(3.1) Theory

1.1) Properties of the Array.

Consider the properties of an artificial dielectric fabricated om conducting sheets in which are cut patterns of resonant ts, the susceptance of each sheet being given by

$$\frac{B_r}{Y} = Q_L\left(\frac{\lambda_r}{\lambda} - \frac{\lambda}{\lambda_r}\right) \quad . \quad . \quad . \quad . \quad . \quad (16)$$

opagation through this array, neglecting the conductivity of e slots, is defined by

$$s\left(\frac{2\pi b}{\lambda_0} n_0\right) = \cos\left(\frac{2\pi b}{\lambda_0}\sqrt{\epsilon_r}\right) - \frac{1}{2}Q_L\left(\frac{\lambda_r}{\lambda} - \frac{\lambda}{\lambda_r}\right)\sin\left(\frac{2\pi b}{\lambda_0}\sqrt{\epsilon_r}\right)$$
$$. \quad . \quad . \quad . \quad . \quad (17)$$

e cut-off wavelengths for propagation in the first pass-band e obtained when $\cos\left(\frac{2\pi b}{\lambda_0} n_0\right) = \pm 1$, so that λ_2, the upper t-off wavelength where $n_0 = 0$, is given by

$$1 = \cos\left(\frac{2\pi b}{\lambda_r}\frac{\lambda_r}{\lambda_2}\right) - \frac{1}{2}Q_L\left(\frac{\lambda_r}{\lambda_2} - \frac{\lambda_2}{\lambda_r}\right)\sin\left(\frac{2\pi b}{\lambda_r}\frac{\lambda_r}{\lambda_2}\right) \quad . \quad (18)$$

e lower cut-off wavelength λ_1 is given by

$$-1 = \cos\left(\frac{2\pi b}{\lambda_r}\frac{\lambda_r}{\lambda_1}\right) - \frac{1}{2}Q_L\left(\frac{\lambda_r}{\lambda_1} - \frac{\lambda_1}{\lambda_r}\right)\sin\left(\frac{2\pi b}{\lambda_r}\frac{\lambda_r}{\lambda_1}\right) \quad . \quad (19)$$

ie cut-off refractive index at $\lambda_0 = \sqrt{\epsilon_r}\lambda_1$, n_c, is given by

$$n_c = \sqrt{\epsilon_r}\frac{\lambda_1}{\lambda_r}\frac{1}{2\frac{b}{\lambda_r}} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

aphs of λ_1/λ_r, λ_2/λ_r and n_c as functions of b/λ_r have been orded elsewhere.[6] An expression for the dispersion is obtained differentiating eqn. (17) with respect to λ_0: this is simplified resonance, where $n_0 = \sqrt{\epsilon_r}$, to give

$$\lambda_0\frac{dn_0}{d\lambda_0} = -\frac{Q_L}{2\pi b/\lambda_r} \text{ at } \lambda = \lambda_r \quad . \quad . \quad . \quad (21)$$

fractive-index characteristics for values of b/λ_r equal to 0·25, 2 and 0·15, with values of Q_L equal to 2, 4 and 8 are shown Fig. 8. The dispersion and n_c increase with increases in Q_L, l as the spacing b is increased, in agreement with a considera- n of the physical mechanism of operation. The refractive- ex characteristics pass through a point of inflection in the nity of resonance.

Z_{ni} and ρ_i are given by eqns. (8) and (6) in the absence of ctive coupling between successive sheets of slots.

2) Closely-spaced Arrangements of Resonant Slots in Conducting Sheets.

The spacings of the slots within each sheet, a and c, must be than $\lambda_0/2\sqrt{\epsilon_r}$ to avoid setting up reflected waves within array, and must be less than $4b/3$ to avoid reactive coupling ween successive sheets. The highest value of Q_L will be ained by using straight narrow half-wave slots in the thinnest

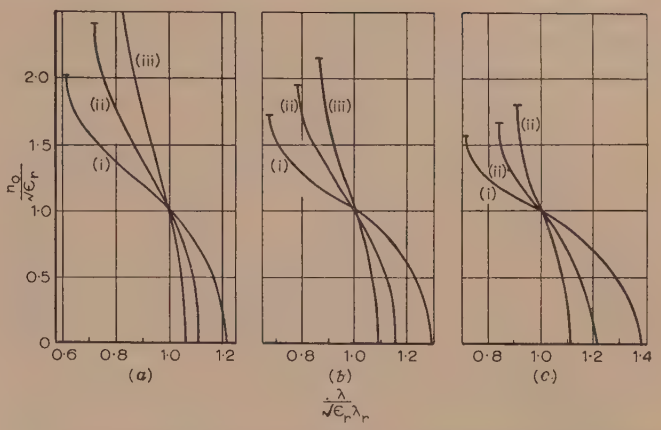


Fig. 8.—Refractive index of the resonant slot dielectric as a function of wavelength.

(i) $Q_L = 2$.
(ii) $Q_L = 4$.
(iii) $Q_L = 8$.
(a) $b/\lambda_r = 0{\cdot}15$.
(b) $b/\lambda_r = 0{\cdot}20$.
(c) $b/\lambda_r = 0{\cdot}25$.

possible sheet. Since the half-wave slot is nearly as long as the largest permissible spacing in the a direction, the slots cannot be perpendicular to the E-vector but must be inclined, symmetry being restored by adopting a pair of crossed slots as the individual element.

The experience of previous workers[2, 9] indicates that the behaviour of a closely-spaced arrangement of resonant elements will differ considerably from that of a single isolated element. An isolated narrow slot is resonant at a wavelength slightly less than twice its length, but an array of the same slots, positioned less than their own length apart, will be resonant at a lower wavelength with a lower value of Q_L. There is no theory to calculate the reductions, so that, in the present case, pilot experiments were carried out before an artificial-dielectric array was designed to be resonant in a particular wavelength range.

The 3 cm parallel-plate transmission line[10] built at Imperial College was convenient to use for pilot experiments. It was envisaged that the resonant sheets for measurements at 1 cm would be produced by photo-etching 0·002 in slot patterns in 0·001 in thick copper foil, this being the narrowest slot which can be photo-etched with accuracy. The slots to be examined at 3 cm wavelength were therefore required to be 0·006 in wide in 0·003 in thick foil. The correct width was consistently obtained by hand-cutting, but the length and angle of slot were not maintained over the large numbers required. Accordingly a design with crossed dumb-bell slots was adopted (see Fig. 9) so that the overall dimensions could be accurately maintained by punching out the dumb-bell circles with a precision punch. Each sheet contained two rows of slots (for support in polystyrene) to ensure that $c < \lambda_0/2\sqrt{\epsilon_r}$, since the spacings between the parallel plates in the spectrometer and transmission line are only slightly less than $\lambda_0/2$.

Test sheets were designed with values of l equal to (1·2, 1·1, 1·0, 0·9, 0·8) $\lambda_0/2\sqrt{\epsilon_r}$, where l is the length of the equivalent straight slot, which, as a rough guide, was assumed to be that with the same perimeter as the dumb-bell slot. The dumb-bell diameter was $\frac{1}{16}$ in. Eight holes were punched simultaneously to locate the ends of four pairs of slots, and these holes were spaced equally to give a symmetrical clearance around the rectangular area occupied by each crossed pair. Subsequent to the design of the test sheets an empirically-based method of determining the resonant wavelength of dumb-bell slots has been described,[11] which has proved quite accurate for single-slot

waveguide irises. The dimensions of the test sheets are given in Table 1, together with the resonant wavelength expected for the isolated slot as calculated by both methods.

Table 1

DIMENSIONS OF DUMB-BELL SLOTS FOR TEST SHEETS

Pattern No.	1	2	3	4	5
l', cm	0·692	0·592	0·492	0·392	0·292
L, cm	1·008	0·908	0·808	0·708	0·608
p, cm	0·785	0·675	0·563	0·444	0·312
a, cm	1·110	1·000	0·888	0·769	0·637
θ	22·5°	25·75°	30°	36·2°	46·2°
Calculated λ_r of single slot, cm:					
Approximate method	2·38	2·18	1·98	1·78	1·58
Empirically based method	2·92	2·66	2·40	2·12	1·83

The arrangement of the slots in each sheet is shown in Fig. 9, the dimensions in the Figure being those for pattern No. 1. Resonance is required near $\lambda_0 = 3{\cdot}20$ cm, i.e. $\lambda_r \simeq 2{\cdot}0$ cm in polystyrene. Fixed dimensions are as follows: $s = 0{\cdot}325$ cm, $c = 0{\cdot}65$ cm, $w = 0{\cdot}015$ cm, $d = 0{\cdot}159$ cm, and the clearance between the rectangular areas occupied by each pair of dumb-bell slots is 0·165 cm.

(3.2) Pilot Experiments at 3 cm

(3.2.1) Manufacture of the Test Sheets.

A combination of punching and engraving processes was used to machine the slot patterns in strips of 0·003 in copper foil, 28 in long. The strips were cut and rolled to an exact 1·000 in width with smooth edges for use as datum lines.

A 1·000 in channel located the strip, while the pattern of eight holes (dimensions shown in Fig. 9) was punched in the

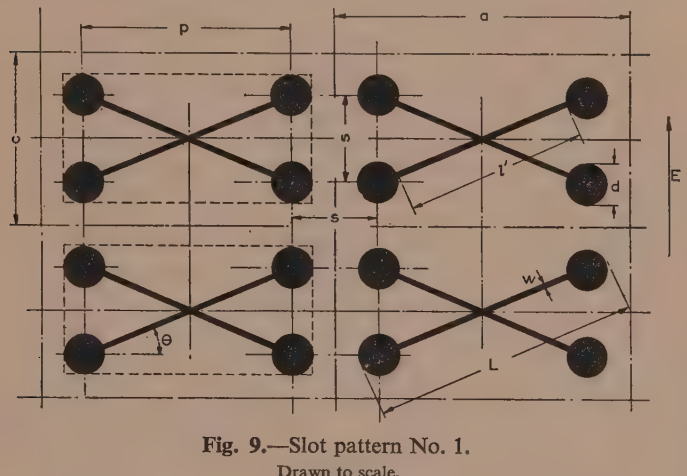


Fig. 9.—Slot pattern No. 1.
Drawn to scale.

centre. The dimension p, which determines the overall length of the slot, was set after each punching operation with a travelling microscope attached to the die block by a calibrated screw thread. The microscope could be set in position to within ±0·001 in.

Slots 0·006 in wide were cut by allowing a sharply-pointed engraving cutter to just break through the foil. Cleanly cut edges were maintained by running the cutter as fast as possible and keeping it continuously sharpened and lubricated with paraffin: the foil was supported on a strip of Paxolin. The engraver followed a pattern extending into the centres of the dumb-bell circles. The strips were finished to 20 in lengths, polished with emergy cloth and bent into channel-section shape so that they could be supported on polystyrene slabs. The final treatment was a short dip in chromic acid to remove oxidatio and any small particles clinging to the slot edges.

(3.2.2) Measurements on Sheets of Slots in the 3 cm Transmission Lin

The test sheets were sandwiched between polystyrene slab 1 cm thick (i.e. approximately $\lambda_r/2$), and the plates of the trans mission line were clamped firmly over the assembly. Standing wave measurements were made in front of the specimens ove a wide range of frequencies, the line being terminated with matching section. The circle diagram was used to extract th admittances of the sheets, which are plotted in Fig. 10(*a*) in th form $Y_r = G_r + jB_r$.

Eqn. (16) can be rearranged in the form:

$$\frac{B_r}{Y\left[1-\left(\frac{f_r}{f}\right)^2\right]} = Q_L\frac{f}{f_r} \quad . \quad . \quad . \quad . \quad (22$$

The validity of eqn. (16) to represent the susceptance of th sheets away from resonance can be examined by plotting th experimental values in the form given in eqn. (22); a graph o this kind is shown in Fig. 10(*b*) for pattern No. 1.

Q_L is evaluated from the experimental results by using

$$Q_L = \frac{f_r}{2}\left(\frac{dB_r}{df}\right)_{f=f_r} \quad . \quad . \quad . \quad . \quad . \quad (23$$

Results taken from Fig. 10(*a*) are shown in Table 2.

Table 2

RESULTS OF TRANSMISSION LINE MEASUREMENTS

Slot pattern	λ_r	Q_L
	cm	
1.	1·98	1·94
2.	1·81 (extrapolated)	2·10
1. (alternate pairs)	2·17	6·5
(one asymmetric row) ..	2·13	7·8
2. (alternate pairs)	1·93	6·2
(one asymmetric row) ..	1·87	6·7

The resonant frequencies of the remaining sheets were too hig to be reached by the range of measurement, showing that thes slots were too short. The effect of coupling between slots i both the *a* and *c* directions was demonstrated by making add tional measurements with one entire row of slots blanked o (leaving a single asymmetric row) and with alternate pairs o slots blanked off (i.e. double *a* spacing). These results are als included in Table 2.

The reductions in Q_L and λ_r which take place with slots whic are closely spaced are strikingly shown in Table 2. Simila effects have been noted in double-slot waveguide irises, an previous workers have found that Q_L for a double-slot iris i reduced by a factor of 3 from that of a single-slot iris when th slots are positioned equidistant between the long wall and th centre-line of the waveguide.[12] Considerable coupling must b responsible for this effect, and it was shown that this couplin was appreciably dependent on the dimensions of the individua slot and the disposition of the slots in the iris. No theoretica treatment has been given.

The conductance of sheet No. 1 was obtained at resonanc from a measurement of the admittance of a slab of five sheet spaced apart by $\lambda_r/2$ of dielectric. The slab was terminate by a $\lambda_r/4$ section of dielectric and a short-circuit to give a effective open-circuit. It was found that $G_r = 0{\cdot}06Y$ at $\lambda = \lambda$

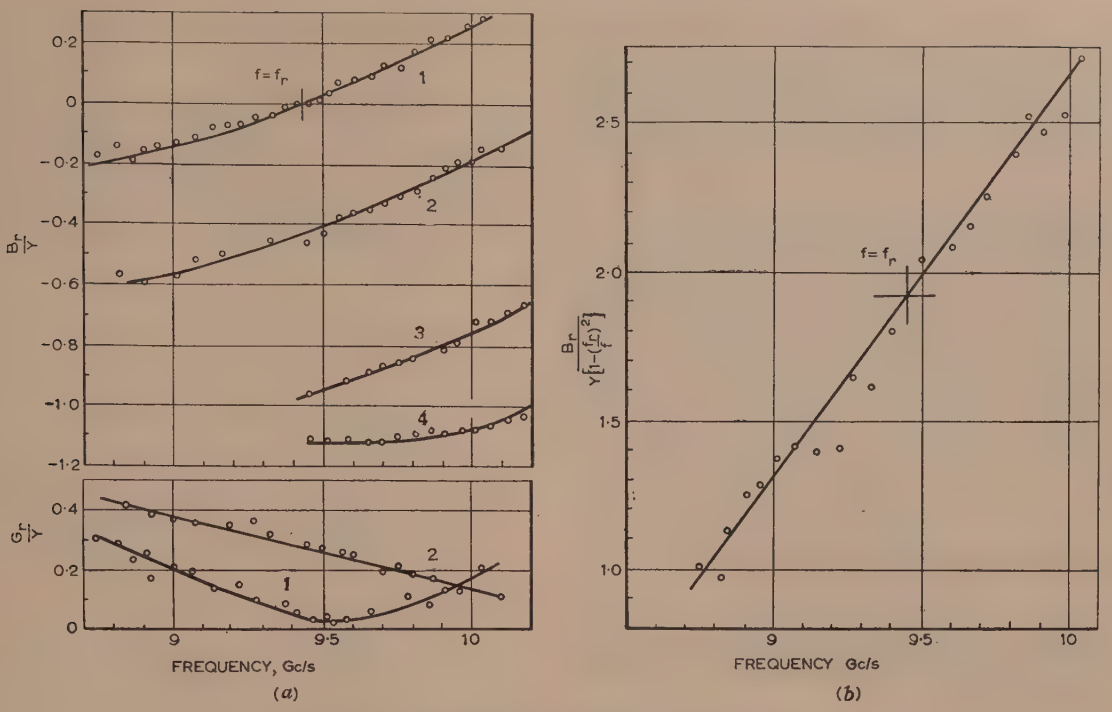


Fig. 10.—Characteristics of test sheets of slots.
(*a*) Measured admittances of test sheets of slots as functions of frequency.
(*b*) $B_r/Y[1-(f_r/f)^2]$ as a function of frequency, Pattern No. 1.

(3.3) Prism Measurements at 1 cm

The dimensions of sheet No. 1 were scaled down by 0·39 to ake a pattern resonant in the vicinity of $\lambda_0 = 1{\cdot}25$ cm. The ngle θ was reduced to 20° because the ratio of the plate spacings the spectrometer and transmission line is 0·36; a thus became navoidably equal to $0{\cdot}55\,\lambda_r$. A ×40 drawing of the pattern as made for the photo-etching process.

The photo-etched sheets were bent into channel-section shape nd cemented with benzene between machined polystyrene abs to make up prisms of the following dimensions:

30° prism, $b = 0{\cdot}353$ cm, $b/\lambda_r = 0{\cdot}45$ requiring 33 sheets of slots.

20° prism, $b = 0{\cdot}159$ cm (standard $\frac{1}{16}$ in sheet), $b/\lambda_r = 0{\cdot}185$ requiring 40 sheets of slots.

Machining difficulties made it impossible to use a spacing naller than $\frac{1}{16}$ in.

Refractive-index and reflection-coefficient measurements were ade as with the rodded prism. Experimental and theoretical aracteristics are compared in Fig. 11. The value 2·5 has en used for Q_L to fit the observations, and this improvement er the figures shown in Table 2 may be ascribed to the superior ethod of manufacture of the sheets for the prisms. (Examina-on under a microscope showed that the photo-etching had oduced slots with very clean-cut edges.) λ_r for the array $\lambda_r = 0{\cdot}45$ is the same as that intended for the isolated sheet, t λ_r for the array $b/\lambda_r = 0{\cdot}185$ is increased from this value. is can be ascribed to the effect of reactive coupling between ccessive sheets, since the ratio of a/b for the array with the naller spacing is such that coupling between the closely-spaced eets is not prohibited.

It was not possible to make measurements with a parallel-led strip of the resonant slot material, because the reflection efficient of the strip increased very rapidly with angle of cidence, becoming as high as 0·8 at $i = 20°$. An approximate lue for the attenuation of the 30° prism was measured at $\lambda = \lambda_r$ by comparing the power transmitted through it with that transmitted through the same path length of a 30° solid polystyrene prism. A value of 0·9 dB/cm was obtained. Using the result obtained in Section 3.2.2 the calculated value is 0·76 dB/cm

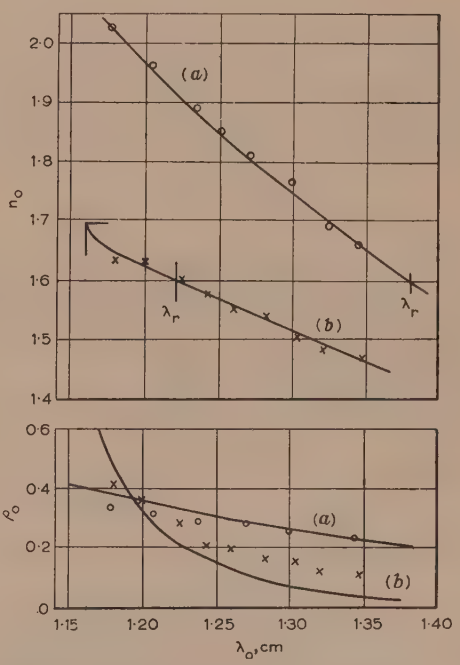


Fig. 11.—Theoretical and experimental characteristics of resonant slot dielectric.
(*a*) $b/\lambda_r = 0{\cdot}185$, $Q_L = 2{\cdot}5$.
(*b*) $b/\lambda_r = 0{\cdot}45$, $Q_L = 2{\cdot}5$.
○ × Measured values.
Theoretical characteristics calculated from eqns. (17) and (8).

for $G_r = 0{\cdot}06\ Y$. The attenuation will rise on either side of resonance following the increases in G_r shown in Fig. 10(*a*).

An attempt to increase Q_L by using sheets with only a single row of slots, centrally placed, was unsuccessful because the resonant wavelength was shifted outside the range of measurement.

(4) DESIGN OF A BEAM-SCANNING PRISM

(4.1) Refraction Produced by a Dispersive Prism

In discussing the refraction produced by a dispersive prism, indicated by eqn. (1), it is convenient to substitute E, the angle of emergence, for the sum of $(P + D)$. D is negative if $n < 1$ and positive if $n > 1$, and although E can take any value between zero and 90°, safety limitations are imposed on the permitted beam swing for practical reasons. The first limitation restricts E to 80°, so that the emergent beam is at least 10° removed from grazing incidence, i.e.

$$\hat{n} = \frac{\sin 80°}{\sin P} \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

The second limitation is arbitrarily imposed as

$$\check{n} = 0{\cdot}1\hat{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

Since E is 5° at $n = \check{n}$ the permitted beam swing within the safety limitations is 75°. A total change in refractive index of $\delta\hat{n}$ is required to produce this swing, and so

$$\delta\hat{n} = \frac{18}{11}\bar{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

and

$$\hat{n} = \frac{20}{11}\bar{n} \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

where $\bar{n}$ is the mean value of n. Eqns. (24), (26) and (27) define the conditions required for 75° beam swing, and are illustrated graphically in Fig. 12(*a*). Eqns. (1) and (24) can be combined to give

$$\sin E = \frac{n}{\hat{n}} \sin 80° \quad . \quad . \quad . \quad . \quad . \quad (28)$$

Eqn. (28) has been used to plot the universal curve shown in Fig. 12(*b*). The prism angle must not be smaller than the values shown in Fig. 12(*a*) or the available beam swing will be considerably restricted because the permissible excursion in n will not cover the upper part of Fig. 12(*b*).

Increments in the angle of emergence, δE, will be produced by increments of refractive index as follows:

$$\delta E = \arcsin\left[\left(\bar{n} + \frac{\delta n}{2}\right)\sin P\right] - \arcsin\left[\left(\bar{n} - \frac{\delta n}{2}\right)\sin P\right] \quad . \quad . \quad . \quad . \quad (29)$$

The greatest permissible value of P for each value of δn will be

$$\arcsin\left[\sin 80° / \left(\bar{n} + \frac{\delta n}{2}\right)\right]$$

so that the largest available beam swing, $\delta\hat{E}$, is given by

$$\delta\hat{E} = 80° - \arcsin\left(\frac{\bar{n} - \frac{\delta n}{2}}{\bar{n} + \frac{\delta n}{2}}\sin 80°\right) \quad . \quad . \quad (30)$$

Values of $\delta\hat{E}$ and δE are plotted in Fig. 13 for three values of $\bar{n}$.

After refraction the aerial aperture is approximately magnified by the factor

$$M = \sqrt{\frac{1 - n^2 \sin^2 P}{1 - \sin^2 P}} \quad . \quad . \quad . \quad . \quad (31)$$

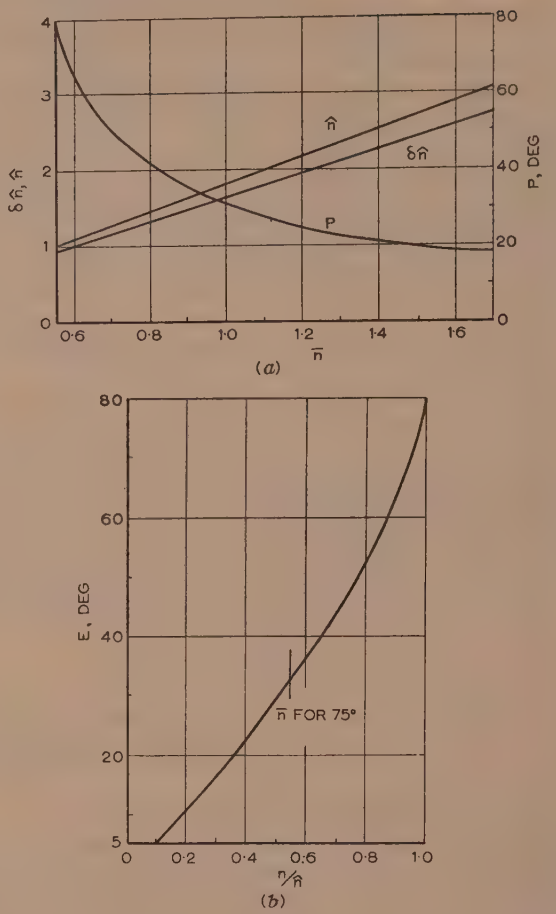


Fig. 12.—General properties of a beam-scanning prism.
(*a*) Conditions required for 75° beam swing.
(*b*) Universal curve of angle of emergence.

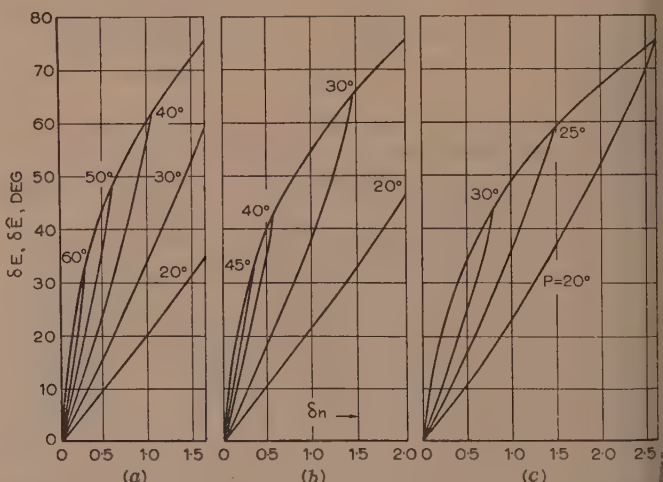


Fig. 13.—Beam swings as functions of δn for various prism angles.
(*a*) $\bar{n} = 1{\cdot}0$.
(*b*) $\bar{n} = 1{\cdot}25$.
(*c*) $\bar{n} = 1{\cdot}60$.

so that the effective aperture is increased if n is less than unity. Eqn. (31) is not accurate, because an aerial does not radiate a parallel beam, but the magnifying effect of the prism must always be considered, since the radiated beam width is inversely proportional to the width of the effective aperture.

(4.2) Use of Artificial Dielectrics in Dispersive Prisms

2.1) General Considerations.

Backward-wave oscillators in present production can provide requency sweep of ±25% with electronically-controlled tuning. ıe availability of this type of oscillator enables a beam-scanning stem using a dispersive prism to be designed.

Increments in wavelength will produce increments in the fractive index of the prism according to the relation

$$\delta n = \lambda_0 \frac{dn}{d\lambda_0} \frac{\delta\lambda_0}{\lambda_0} \quad . \quad . \quad . \quad . \quad . \quad (32)$$

ıere $\lambda_0 \frac{dn}{d\lambda_0}$ is the dispersion of the artificial dielectric.

To achieve a 75° beam swing, eqns. (26) and (27) show that r the rodded dielectric $\hat{n} = 0{\cdot}9\sqrt{\epsilon_r}$ and $\bar{n} = 0{\cdot}495\sqrt{\epsilon_r}$, and r the resonant slot dielectric $\hat{n} = 1{\cdot}82\sqrt{\epsilon_r}$ when $\bar{n} = \sqrt{\epsilon_r}$. The hypotenuse of an artificial-dielectric prism will behave as a ffraction grating of spacing $b/\sin P$, so that the condition for oiding the production of side waves at this interface is[8]

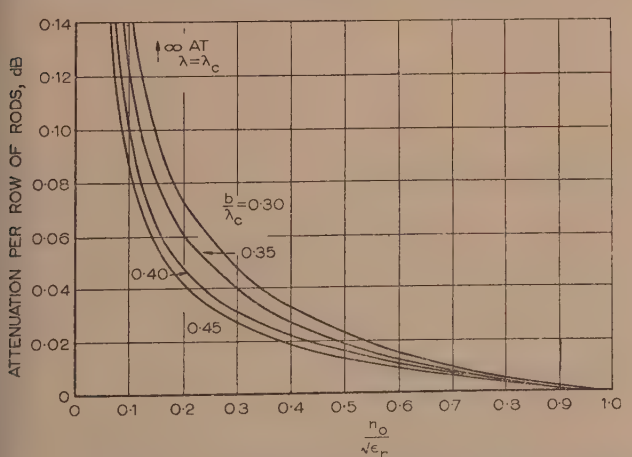
$$\frac{\lambda}{b} - 1 \geqslant \operatorname{cosec} P \quad . \quad . \quad . \quad . \quad . \quad (33)$$

ıis relation shows that λ/b should not be less than 2·15 in a ° prism, and not less than 2·4 in a 45° prism, if side waves e to be avoided. The side waves represent a loss of power d a possible source of interference, and although these effects ll be small when the side waves are scattered at small angles, it ould be noted that the wave which travels back into the prism ay affect the matching of the first surface.

2.2) Rodded Dielectric.

Beam-swing characteristics for the rodded dielectric can be epared from Figs. 3 and 12(*b*). The wavelength sweep required r 75° scan is reduced as b/λ_c is increased and the beam swing aracteristic becomes more non-linear. The scanning per- rmance of the rodded dielectric is assessed in Table 3.

Universal attenuation characteristics for the arrays in which = *b* can be prepared from eqn. (14) using the simplified value X_1, as in Fig. 14. These curves are inaccurate when eqn. (3) used to define propagation through the array, as for large lues of b/λ_c. The attenuation of arrays in which *r* and *a* ve been reduced to make X_2/X_1 a negligible quantity will be creased from the values shown in Fig. 14.



g. 14.—Attenuation characteristics for those rodded arrays where *a* = *b*, calculated for copper rods, supported in polystyrene, at a frequency of 27 Gc/s.

(4.2.3) Resonant Slot Dielectric.

Beam-swing characteristics for this artificia dielectric can be prepared from Figs. 8 and 12(*b*). The dispersion is increased by decreasing b/λ_r, and the dispersion at any value of b/λ_r is increased if Q_L can be increased. Q_L is reduced by reducing the spacings between the slots in each sheet and the spacings between the sheets, so that there is a limiting value to the reduction of *b* beyond which the dispersion cannot be increased. This point was not reached in the experiments described in Section 3.3.

The restrictions on the experimental work made it impossible to investigate a range of slot configurations, and in the absence of an established theory, an empirical approach is required to the design of an array with a low value of b/λ_r. The following steps are suggested to improve the overall value of Q_L:

(*a*) The slots could be inclined at a large angle so that the coupling with the *E*-vector was reduced.
(*b*) A diagonal arrangement of slots could be used in each sheet and the sheets staggered to minimize the coupling between adjacent slots and between successive sheets.
(*c*) Dumb-bell slots should be avoided, since the individual value of Q_L is low and the dumb-bell ends may allow extra coupling.

It is impossible to predict the attenuation of the resonant array, and measurement has only been possible at $\lambda = \lambda_r$. The following steps should be taken to reduce the unavoidable losses in each sheet:

(i) The slot width and sheet thickness should be as small as possible.
(ii) The highest r.f. surface conductivity should be achieved.
(iii) The slots should be cut very cleanly.

The scanning performances of the rodded and resonant-slot dielectrics are compared in Table 3.

Table 3

PERFORMANCES OF THE RODDED AND RESONANT-SLOT DIELECTRICS IN DISPERSIVE PRISMS

RODDED DIELECTRIC		RESONANT-SLOT DIELECTRIC
	75° Beam Scanning	
b/λ_c	Wavelength sweep	$b/\lambda_r = 0{\cdot}15$, $Q_L = 2{\cdot}5$,
0·45	1·1 : 1	Wavelength sweep $= \lambda_r + 20\%, -27\%$
0·40	1·25 : 1	
0·35	1·4 : 1	$b/\lambda_r = 0{\cdot}15$, $Q_L = 4{\cdot}0$,
0·30	1·65 : 1	Wavelength sweep $= \lambda_r + 11\%, -12\%$
	Reflection Losses	
ρ is increased as b/λ_c is increased and reduced by increasing ϵ_r.		ρ passes through zero at a value of $n \leqslant \sqrt{\epsilon_r}$, and the variation of ρ is made increasingly asymmetrical by increasing ϵ_r.
	Attenuation Losses	
The attenuation is decreased by increasing b/λ_c. With a 45° prism of aperture 20 λ_0, the greatest attenuation along the axis should not exceed 2·8 dB.		The minimum attenuation, at $\lambda = \lambda_r$, is increased by decreasing b/λ_r. The attenuation at $\lambda = \lambda_r$ should not exceed 1·5 dB for a 45° prism, aperture 20 λ_0.
	Other Features	
(*a*) Side waves will be produced over considerable portions of the scan for large values of b/λ_c.		No side waves will be produced in a 30° prism when $b/\lambda_r < 0{\cdot}25$.
(*b*) Larger prism angles will be required since $\bar{n} = 0{\cdot}495\sqrt{\epsilon_r}$.		$\bar{n} = \sqrt{\epsilon_r}$.
(*c*) Dispersive theory is reasonably exact.		Q_L and λ_r, and hence the dispersion, have to be determined empirically.

Reflection losses at the surfaces of dispersive materials can be substantially reduced by a quarter-wave matching technique which is applicable to prisms using either array. This technique is described elsewhere[13] and has been successfully applied to the rodded prism described in Section 2.2.

(5) CONCLUSIONS

Assessment of Table 3 shows that a prism using either the rodded or resonant-slot dielectric will not be suitable for wide-angle scanning unless a considerable change in wavelength can be tolerated. The available beam swing could easily be doubled by using a pair of prisms producing deviations in opposite senses fed by a suitable microwave switch. It is suggested that a high-speed scanning system using either artificial dielectric could be designed to scan through 20°, which would avoid the principal disadvantages noted here. Such a system would be of use for roll-stabilizing ship-borne equipment or for search radar in aircraft and high-speed missiles. Calculations have shown that the scanning performance of the dispersive prism compares favourably with that of a parallel-sided slab which has a distribution of dispersive properties across its aperture,[6] and also with that of a linear array of radiators fed by dispersive waveguide.[14]

(6) ACKNOWLEDGMENTS

The work described is part of the programme of the Radio Research Board and is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

Acknowledgment is made to Prof. A. Tustin for facilities provided in the Electrical Engineering Department, Imperial College, where the work was carried out.

Acknowledgment is made of the kind provision of 1 cm klystron oscillators by the General Electric Company Limited, the National Physical Laboratory and the Radio Research Station. The jig-boring of the rodded material described in Section 2 was generously undertaken by Metropolitan-Vickers Electrical Co., Ltd.

(7) REFERENCES

(1) Brown, J., and Jackson, Willis: 'The Properties of Art ficial Dielectrics at Centimetre Wavelengths', *Proceedin I.E.E.*, Paper No. 1699 R, January, 1955 (**102** B, p. 11)

(2) El-Kharadly, M. M. Z.: 'Some Experiments on Artifici Dielectrics at Centimetre Wavelengths', *ibid.*, Pap No. 1700 R, January, 1955 (**102** B, p. 17).

(3) Brown, J.: 'Artificial Dielectrics having Refractive Indic Less than Unity', *ibid.*, Monograph No. 62 R, May, 19 (**100**, Part IV, p. 51).

(4) Marcuvitz, N.: 'Waveguide Handbook', Radiatio Laboratory Series, Vol. 10 (McGraw-Hill, 1951).

(5) Sollom, P. H., and Brown, J.: 'A Centimetre-Wav Parallel-Plate Spectrometer', *Proceedings I.E.E.*, Pap No. 2008 R, May, 1956 (**103** B, p. 419).

(6) Seeley, J. S.: 'An Investigation of the Properties of Di persive Artificial Dielectrics' (University of Londo Ph.D. Thesis, 1957).

(7) Seeley, J. S.: 'A Spectrometer Method for Measuring th Electrical Constants of Lossy Materials', *Proceedin I.E.E.*, Monograph No. 248 R, July, 1957 (**105** C p. 18).

(8) Carne, A.: 'Design Theory of the Rodded Artifici Dielectric', M.Sc. Thesis, University of London, 1957.

(9) Primich, R. I.: 'A Study of the Properties of Artifici Dielectrics' (University of London, Ph.D. Thesis, 1954).

(10) Adey, A. W.: 'Microwave Diffraction by Cylindric Obstacles' (University of London, Ph.D. Thesis, 1954).

(11) Fernando, W. C. L.: 'Resonant Dumb-bell Irises in Wav guide' (Imperial College, D.I.C. Thesis, 1956).

(12) Reginold, I., Carter, J. L., and Garoff, K.: 'Single- an Multi-Iris Resonant Structures', *Proceedings of th Institute of Radio Engineers*, 1952, **40**, p. 861.

(13) Seeley, J. S.: 'The Quarter-Wave Matching of Dispersiv Materials' (see next page).

(14) Smith, J. M.: Unpublished work, Metropolitan-Vicker Electrical Co., Ltd.

[The discussion on the above paper will be found on page 114.]

1.396.677.852 : 621.396.965.4

The Institution of Electrical Engineers
Paper No. 2736 R
Nov. 1958
©

THE QUARTER-WAVE MATCHING OF DISPERSIVE MATERIALS

By J. S. SEELEY, B.Sc.(Eng.), Ph.D., Graduate.

(*The paper was first received* 12*th March, and in revised form* 15*th July*, 1958. *It was published in November*, 1958, *and was read before the* RADIO AND TELECOMMUNICATION SECTION 24*th November*, 1958.)

SUMMARY

Reflections from the surfaces of dispersive materials used in broad-nd aerial systems are highly frequency-dependent. A technique for atching such materials is described, and results are included of a ccessful application to the input surface of a dispersive prism.

LIST OF PRINCIPAL SYMBOLS

X = Reactance of matching grating.
Z = Wave impedance of supporting dielectric material.
Z_0 = Free-space wave impedance.
Z_{n0} = Wave impedance of dispersive material at normal incidence.
ϵ_r = Relative permittivity of supporting dielectric.
λ_0 = Free-space wavelength.
ρ_0 = Modulus of reflection coefficient of dispersive material at normal incidence.
θ_1, θ_2, etc. = Equivalent electrical length of layers of supporting dielectric $= \frac{2\pi\sqrt{\epsilon_r}}{\lambda_0} \times$ (layer thickness).

The symbols in bold face are complex.

(1) INTRODUCTION

Reflections from the surfaces of a dispersive prism used for e scanning of radar beams[1] will lead to a wastage of power nd increases in the input standing-wave ratio and radiated side-be level. Materials for such a prism are required to be highly spersive, so that the reflections are unavoidably frequency-pendent. It will be assumed throughout the paper that the mpedance of these materials is wholly real and that reactive eld effects at the surfaces[2] can be neglected.

The impedance of the prism materials, Z_{n0}, will not be equal the free space impedance Z_0, so a matching layer must mulate the quarter-wave transforming section defined by

$$Z_m^2 = Z_0 Z_{n0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\theta_m = (2q+1)\frac{\pi}{2} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

here Z_m is the matching impedance required and θ_m is the uivalent electrical thickness of the section. The matching yer can be of three types:

(*a*) A quarter-wavelength of solid dielectric, $\epsilon_r = 1/Z_0 Z_{n0}$.
(*b*) A layer of solid dielectric slotted to simulate a quarter-wave transforming layer.
(*c*) A section comprising a reactive grating and lengths of solid dielectric which simulates a quarter-wave transforming layer.

Type (*c*) is adopted in the present case, so that the frequency pendence of the grating can be turned to an advantage in atching an impedance which is frequency-dependent.

The use of a matching grating positioned on the free-space side of the interface has been suggested for use with artificial dielectrics.[3] When the artificial-dielectric array is supported in a solid dielectric it is convenient to support the matching grating in the same material, removing the physical interface to the free space side of the matching section. The arrangement thus becomes similar to the use of a reactive grating to match the surface of a solid dielectric lens,[4] with the addition of a second layer of dielectric to complete a quarter-wave layer.

(2) DESIGN OF A QUARTER-WAVE MATCHING SECTION

The equivalent transmission-line circuit of the matching section is shown in Fig. 1. Design equations for the section are obtained

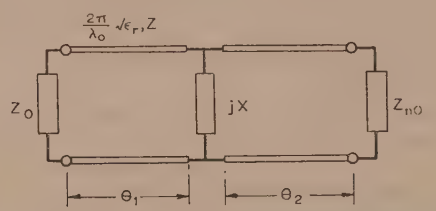


Fig. 1.—Equivalent circuit of matching section.

by equating the transfer matrix of this circuit with that of the quarter-wave transformer, so that

$$\begin{bmatrix} \cos\theta_1 & jZ\sin\theta_1 \\ \frac{j}{Z}\sin\theta_1 & \cos\theta_1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1/jX & 1 \end{bmatrix}\begin{bmatrix} \cos\theta_2 & jZ\sin\theta_2 \\ \frac{j}{Z}\sin\theta_2 & \cos\theta_2 \end{bmatrix}$$

$$\equiv \begin{bmatrix} \cos\theta_m & jZ_m\sin\theta_m \\ \frac{j}{Z_m}\sin\theta_m & \cos\theta_m \end{bmatrix} \quad (3)$$

Since the quarter-wave section is symmetrical:

$$\cos(\theta_1+\theta_2) + \frac{Z}{X}\sin\theta_1\cos\theta_2 = \cos(\theta_1+\theta_2) + \frac{Z}{X}\cos\theta_1\sin\theta_2 \quad (4)$$

i.e. $$\theta_1 = \theta_2 = \theta, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The thickness of the matching layer is given by $d = \frac{\lambda_0\theta}{\pi\sqrt{\epsilon_r}}$

The design equations obtained from the further expansion of eqn. (3) are now expressed as

$$\tan 2\theta = -2\frac{X}{Z} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and

$$\left.\begin{aligned} \tan^2\theta = \frac{Z_m^2}{Z^2} &= \frac{Z_0 Z_{n0}}{Z^2} \\ &= \epsilon_r\frac{Z_{n0}}{Z_0} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (7)$$

Dr. Seeley was formerly in the Electrical Engineering Department, Imperial College Science and Technology, University of London and is now at the Royal Aircraft tablishment.

Eqns. (6) and (7) can be combined in an expression which is more convenient for balancing the frequency dependence of Z_{n0} by a suitable frequency dependence of X/Z:

$$\frac{X}{Z} = \pm \frac{(\epsilon_r Z_{n0} Z_0)^{1/2}}{(\epsilon_r Z_{n0} - Z_0)} \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Eqn. (8) shows that a particular value of Z_{n0} can be matched with either positive or negative reactive gratings. The matching of solid dielectrics and non-dispersive artificial dielectrics having $Z_{n0} > Z_0/\epsilon_r$ is accomplished either with θ slightly greater than $\pi/4$ using an inductive grating, or with θ slightly smaller than $3\pi/4$ using a capacitive grating. The matching of dispersive materials will require a capacitive grating (for which X increases with increase in wavelength) if Z_{n0} decreases with increase in wavelength, and an inductive grating if Z_{n0} increases with increase in wavelength. The dimensions of gratings to give prescribed values of X/Z can be obtained from the tabulations of Marcuvitz[5] and Cohn.[6]

(3) MATCHING THE INPUT SURFACE OF A DISPERSIVE PRISM

The matching technique has been applied to the input surface of a 45° beam-scanning prism, which contains a highly dispersive array of rods.[1] The measured impedance of the prism material at normal incidence is shown in Fig. 2, together with the desired wavelength dependence of X/Z, calculated from eqn. (8). Curves (*a*) and (*b*) show the wavelength dependence of two inductive gratings designed to give an exact match at different wavelengths. Curve (*c*) shows the wavelength dependence of a capacitive grating, which is seen to have the incorrect sense of change with change in wavelength and is not further considered. The wavelength at which to make the exact match is determined by the behaviour of the reflection coefficient ρ and input admittance Y_{in} of the matched surface within the operating range of wavelength.

The input admittance and modulus of the reflection coefficient at normal incidence are given by

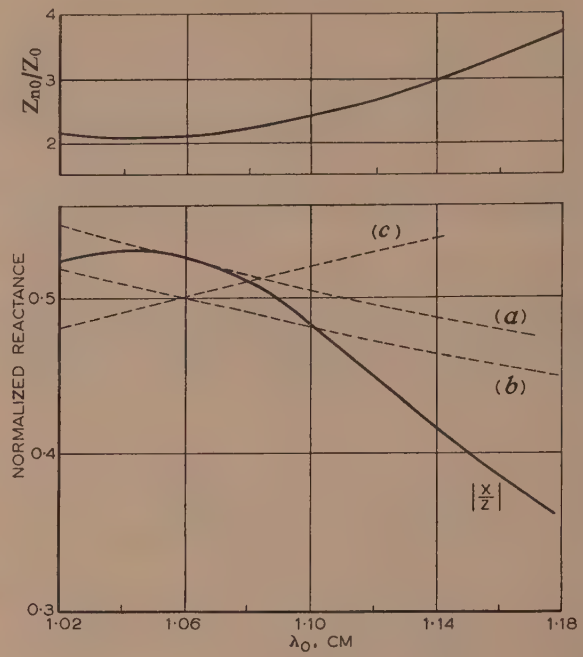


Fig. 2.—Z_{n0} of rodded dielectric and $\left|\frac{X}{Z}\right|$ required for matching.

(*a*) $\frac{X}{Z} = \frac{0{\cdot}5565}{\lambda_0}$, $\theta = 0{\cdot}371\pi$,

(*b*) $\frac{X}{Z} = \frac{0{\cdot}530}{\lambda_0}$, $\theta = 0{\cdot}378\pi$,

(*c*) $\frac{X}{Z} = -0{\cdot}472\lambda_0$, $\theta = 0{\cdot}623\pi$.

$$Y_{in} = G_{in} + jB_{in} = \frac{\epsilon_r}{Z_{n0}} \left\{ \frac{\left[\sqrt{\epsilon_r}(1 - \tan^2\theta) + \frac{Z}{X}\tan\theta\right]^2 + \tan\theta\left(2\sqrt{\epsilon_r}\tan\theta - \frac{Z}{X}\right)\left(2\sqrt{\epsilon_r} + \frac{Z}{X}\tan\theta\right)}{\epsilon_r\left[\sqrt{\epsilon_r}(1 - \tan^2\theta) + \frac{Z}{X}\tan\theta\right]^2 + \left(\frac{Z_0}{Z_{n0}}\tan\theta\right)^2\left(2\sqrt{\epsilon_r} + \frac{Z}{X}\tan\theta\right)^2} \right\}$$

$$+ j\sqrt{\epsilon_r} \frac{\left[\epsilon_r\left(2\sqrt{\epsilon_r}\tan\theta - \frac{Z}{X}\right) - \left(\frac{Z_0}{Z_{n0}}\right)^2\tan\theta\left(2\sqrt{\epsilon_r} + \frac{Z}{X}\tan\theta\right)\right]\left[\sqrt{\epsilon_r}(1 - \tan^2\theta) + \frac{Z}{X}\tan\theta\right]}{\epsilon_r\left[\sqrt{\epsilon_r}(1 - \tan^2\theta) + \frac{Z}{X}\tan\theta\right]^2 + \left(\frac{Z_0}{Z_{n0}}\tan\theta\right)^2\left(2\sqrt{\epsilon_r} + \frac{Z}{X}\tan\theta\right)^2} \quad . \quad (9$$

$$\rho_0^2 = \frac{\left\{\sqrt{\epsilon_r}\left(\frac{Z_0}{Z_{n0}} - 1\right)\left[\sqrt{\epsilon_r}(1 - \tan^2\theta) + \frac{Z}{X}\tan\theta\right]\right\}^2 + \left[\sqrt{\epsilon_r}\left(2\sqrt{\epsilon_r}\tan\theta - \frac{Z}{X}\right) - \frac{Z_0}{Z_{n0}}\tan\theta\left(2\sqrt{\epsilon_r} + \frac{Z}{X}\tan\theta\right)\right]^2}{\left\{\sqrt{\epsilon_r}\left(\frac{Z_0}{Z_{n0}} + 1\right)\left[\sqrt{\epsilon_r}(1 - \tan^2\theta) + \frac{Z}{X}\tan\theta\right]\right\}^2 + \left[\sqrt{\epsilon_r}\left(2\sqrt{\epsilon_r}\tan\theta - \frac{Z}{X}\right) + \frac{Z_0}{Z_{n0}}\tan\theta\left(2\sqrt{\epsilon_r} + \frac{Z}{X}\tan\theta\right)\right]^2} \quad (10$$

These expressions are too unwieldy to determine the best value of X/Z in the general case, so that, for a particular matching problem, it will be necessary to compare the values of Y_{in} and ρ_0 computed for various values of X/Z. Computed curves for the rodded dielectric when matched with layers based on gratings (*a*) and (*b*) are shown in Fig. 3. A row of rods perpendicular to the direction of propagation and aligned parallel to the incident electric field forms a convenient inductive grating, the reactance of which is given by

$$\frac{X}{Z} \simeq \sqrt{\epsilon_r}\frac{a}{\lambda_0}\log_e\left(\frac{a}{2\pi r}\right) \qquad \text{for thin rods.}$$

Design (*b*) was adopted for practical examination because copper wire of a convenient standard size fitted the required dimensions. The arrangement of the matching layer is shown in Fig. 4. The supporting dielectric was polystyrene, and the difficulty of drilling very small holes in thin strips of this material was such that only one surface of the prism was matched.

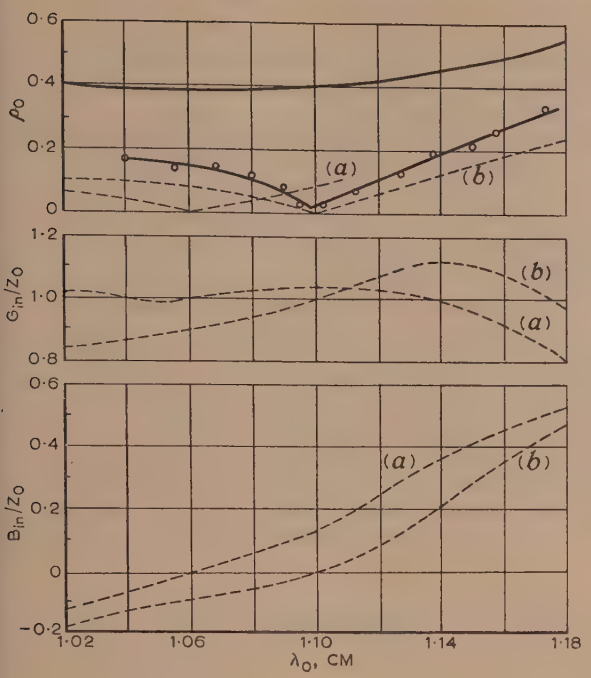


Fig. 3.—Performance of matching sections applied to rodded dielectric.

(*a*) and (*b*) Sections as defined in Fig. 2.
——— Rodded dielectric, not matched.
○ ○ / ○ Rodded dielectric, matched with section (*b*).
- - - - - - Theoretical behaviour.

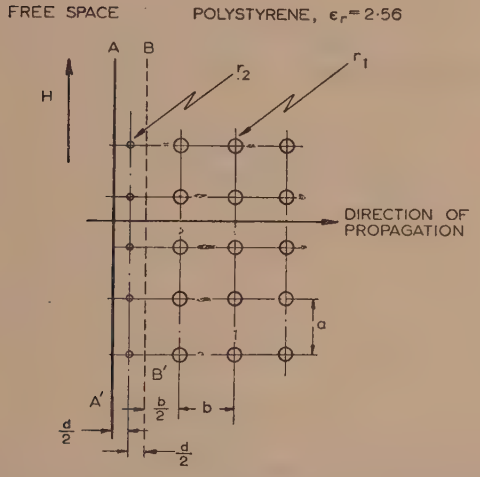


Fig. 4.—Arrangement of matching layer for rodded dielectric.

$a = b = 0{\cdot}118$ in, $d = 0{\cdot}102$ in.
$2r_1 = 0{\cdot}024$ in (No. 23 s.w.g.), $2r_2 = 0{\cdot}0124$ in (No. 30 s.w.g.).
AA′ = Physical interface.
BB′ = Effective interface of rodded dielectric.

The reflection coefficient of the matched surface was measured normal incidence in a parallel-plate spectrometer.[1] Experintal values are shown in Fig. 3, and they are in good agreement h the theoretical predictions. Small discrepancies may be e to the attenuation of the composite prism material and the ctive fields at the surface, and also to the use of an approxite expression for the reactance of the matching grating.

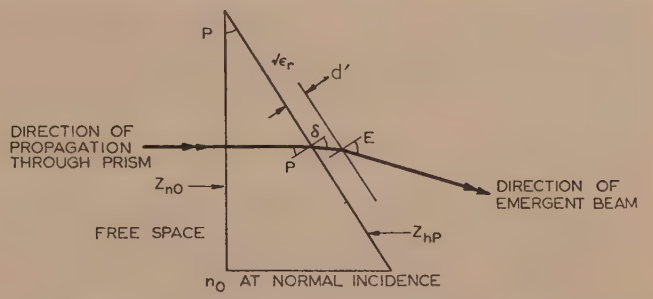


Fig. 5.—Matching layer for hypotenuse of prism.
Not to scale.

(4) MATCHING AT OBLIQUE INCIDENCE

The matching of the refracting surface (i.e. hypotenuse) of a beam-scanning prism is complicated by large variations in the angle of emergence of the beam from this surface.

The beam emerging from the prism will be refracted through an angle δ by a layer of plain dielectric at the hypotenuse (see Fig. 5), where

$$\delta = \arcsin\left(\frac{n_0 \sin P}{\sqrt{\epsilon_r}}\right) \quad . \quad . \quad . \quad . \quad (11)$$

The matching eqns. (6) and (7) become, at oblique incidence,

$$\tan 2\theta' = -2\frac{X'}{Z} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

and

$$\tan^2 \theta' = \frac{\epsilon_r \sec (E) Z_{hP}}{Z_0 \sec^2 \delta} \quad . \quad . \quad . \quad . \quad (13)$$

where $E = \arcsin (n_0 \sin P)$.

X'/Z = Reactance required for matching at angle δ.
Z_{hP} = Impedance of the hypotenuse surface of the prism at the angle of incidence P.

The thickness of the matching layer is given by

$$d' = \frac{\lambda_0 \theta'}{\pi\sqrt{(\epsilon_r - \sin^2 P)}}.$$

No practical examination of the matching of a dispersive refracting surface has been made. Computation of Y_{in} and ρ_{hP} will only be of interest for a particular case, so that general comments alone will be made here. It will be highly desirable to match the refracting surface, since, in general, $\rho_{hP} > \rho_0$ and a wave reflected back into the prism by the hypotenuse surface may affect the matching of the first surface. It is expected that the greater variations in Z_{hP} throughout the operating range of λ_0, allied with variations in E, will result in the matched performance being inferior to that of a surface used only at normal incidence.

(5) CONCLUSIONS

The validity of the technique described for matching dispersive materials has been established by experiment. In a particular material values of ρ_0 equal to 0·39 (at $\lambda_0 = 1{\cdot}04$ cm), 0·41 (at 1·10 cm), 0·5 (at 1·16 cm) have been reduced to 0·16, 0·03 and 0·26 by matching. It is expected that an improvement in matching and an increase in bandwidth could be obtained by using a cascade of two or more quarter-wave sections designed to have a maximally flat response.

(6) ACKNOWLEDGMENTS

The work described is part of the programme of the Radio Research Board, and the paper is published by permission of the Director of Radio Research of the Department of Scientific and

Industrial Research. The author is grateful to Dr. John Brown, University College, London, for supervision during the course of the work, and to Prof. A. Tustin for facilities provided in the Electrical Engineering Department, Imperial College of Science and Technology.

(7) REFERENCES

(1) SEELEY, J. S., and BROWN, J.: 'The Use of Dispersive Artificial Dielectrics in a Beam-Scanning Prism' (see page 93).

(2) BROWN, J., and SEELEY, J. S.: 'The Fields associated with an Interface between Free Space and an Artificial Dielectric', *ibid.*, Monograph No. 300 R, April, 1958 (**105** C p. 465).

(3) BROWN, J.: 'Artificial Dielectrics having Refractive Indice less than Unity', *ibid.*, Monograph No. 62 R, May, 195 (**100**, Part IV, p. 51).

(4) JONES, E. M. T., and COHN, S. B.: 'Surface Matching o Dielectric Lenses', *Journal of Applied Physics*, 1955, **2**(p. 452.

(5) MARCUVITZ, N.: 'Waveguide Handbook' (McGraw-Hil 1951).

(6) COHN, S. B.: 'Electrolytic-Tank Measurements for Micro wave Delay-Lens Media', *Journal of Applied Physics*, 195(**21**, p. 674.

[The discussion on the above paper will be found on page 114.]

.371 : 621.396.677.859

The Institution of Electrical Engineers
Paper No. 2742 R
Nov. 1958
©

HEORY OF REFLECTIONS FROM THE RODDED-TYPE ARTIFICIAL DIELECTRIC

By A. CARNE, B.A., M.Sc., and J. BROWN, M.A., Ph.D., Associate Member.

(*The paper was first received 4th June, and in revised form 1st August*, 1958. *It was published in November*, 1958, *and was read before the* RADIO AND TELECOMMUNICATION SECTION *24th November*, 1958.)

SUMMARY

t is demonstrated that a composite material of a solid dielectric h embedded conductors can give a wave impedance equivalent to t of free space, and thus eliminate interface reflections. A par-ılar material, the rodded dielectric, is analysed, and design proce-es are given which are exemplified by numerical results. An erimental verification of the design theory when the electric-field tor is wholly parallel to the axes of the rods is described, and the ults show good agreement with the theory. Finally, methods of nufacture, and the tolerances required, are discussed.

LIST OF SYMBOLS

x, y, z = Rectangular axes, z being the direction of normal propagation.
ϵ_r = Relative permittivity of supporting dielectric.
μ_r = Relative permeability of supporting dielectric.
b = Spacing between adjacent rods.
r = Radius of rod.
Z_0 = Wave impedance for plane wave in free space.
Z = Wave impedance for medium other than free space.
λ_0 = Free-space wavelength.
λ = Wavelength in supporting dielectric.
i = Angle of incidence.
i' = Angle of refraction in supporting dielectric.
ρ = Amplitude of reflection coefficient.
$|\rho_m|$ = Modulus of a maximum allowable reflection coefficient.
$\alpha = (1 - |\rho_m|)/(1 + |\rho_m|)$.
β = Phase-change coefficient $= 2\pi/\lambda$.
β_z = z-component of phase-change coefficient.
$\phi = \beta b$.
$A = \pi/\log_e (b/2\pi r)$.
$\zeta = \sqrt{\epsilon_r} Z(\beta, 0)/Z_0$.
$\zeta_m = \epsilon_r \alpha^2$.
$\zeta_1 = Z_0/\alpha$.

e rationalized M.K.S. system of units is used throughout the per.

(1) INTRODUCTION

In several microwave components, such as radomes, it is cessary to have a material which is 'invisible' to the micro-ve radiation. Such a material would not reflect any incident liation or cause any change in the direction of propagation. is readily demonstrated that a material is invisible in this se only if its permittivity and permeability have the same ues as those for free space. A modified material which does introduce any reflection is, however, useful in many cases, ce the resulting refraction can either be corrected or allowed in the design of the associated components. The sandwich e of materials which are widely used in radome construction[1] designed with this restricted requirement, the dimensions of the layers being selected to minimize reflection over as wide a range of frequency and direction of incidence as possible. The reduction in reflection is achieved by a cancellation of the individual reflections which arise at each interface between two adjacent layers. The material described in the paper represents an alternative solution for which the material is reasonably homogeneous and of the same impedance as free space.

When a plane wave is normally incident on the interface between free space and a homogeneous material, partial reflection occurs and the amplitude reflection coefficient is

$$\rho = (Z - Z_0)/(Z + Z_0) \quad . \quad . \quad . \quad . \quad (1)$$

Z and Z_0 being, respectively, the wave impedance for plane waves in the material and in free space. These impedances satisfy the relation

$$Z = Z_0(\mu_r/\epsilon_r)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where μ_r and ϵ_r are, respectively, the relative permeability and permittivity of the material. The reflection coefficient ρ is zero only if Z equals Z_0, and from eqn. (2) it is seen that this implies that μ_r should equal ϵ_r. With one exception, all materials which might be used for radome construction have a relative permeability of unity: even with the exception—the class of ceramic materials known as ferrites—relative permeabilities very different from unity occur only in the presence of strong steady magnetic fields. Accordingly, the relative permittivity must also be almost unity if small reflections are to be obtained: this happens only for foamed materials as used in sandwich-type radomes and which are insufficiently strong mechanically to be used alone. The conclusion is that no ordinary material is likely to have the required property.

There remains the possibility of combining a solid dielectric material with an array of conducting elements to form an artificial dielectric. It has been established that such arrays can be designed to give a very wide range of relative permittivities, including those less than unity.[2] The solid material provides the necessary mechanical strength and the loading by the conductors can be controlled to satisfy the condition of small reflection over as wide a band as possible. The nature of this loading can be considered in general terms by using the similarity between plane-wave propagation in a dielectric and wave propagation on a transmission line.[3] The basic problem is to increase the impedance from the value for the solid material as given by eqn. (2) to the value Z_0 for the composite medium. The impedance of a transmission line can be increased either by increasing the series inductance per unit length or by decreasing the shunt capacitance per unit length. Over a small frequency region, the latter can be achieved by connecting a set of inductances in shunt across the line at regular intervals. The conductors introduced into the solid material must play a corresponding part. Several methods of arranging conducting elements to produce the effect of a shunt inductance are known, namely

(*a*) A regular or completely random array of thin conductors of lengths a little greater than half the wavelength of operation, as measured in the solid dielectric.

r. Carne was formerly with the Metropolitan-Vickers Electrical Co., Ltd., and is at the Services Electronics Research Laboratory.
r. Brown is Reader in Electrical Engineering, University College, London.

(*b*) A succession of parallel conducting plates each pierced by a lattice of small holes.

(*c*) An array of thin conducting wires parallel to the electric field of the incident wave.

The last method is the one which has been studied as it is capable of relatively easy analysis and is easier to realize than the others.

(2) CALCULATION OF THE REFLECTION COEFFICIENT

(2.1) Outline of the Procedure

The composite material being examined consists of a solid dielectric of relative permittivity ϵ_r, in which are embedded wires of radius r regularly spaced at intervals b, as in Fig. 1.

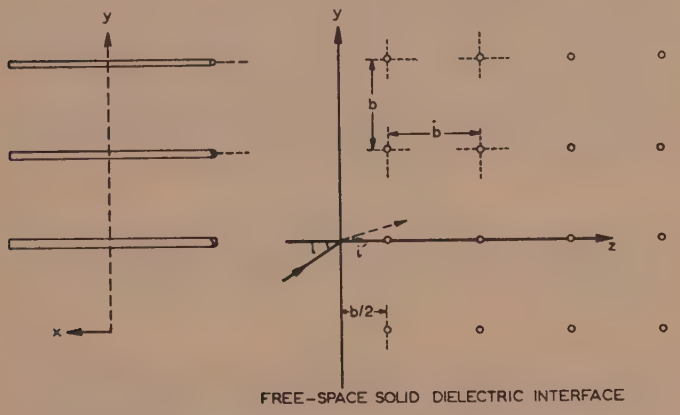


Fig. 1.—General arrangement of rodded dielectric.

The surface of the material is arranged to be at a distance $b/2$ from the plane through the axes of the first set of wires. The reasons for this choice are discussed in Reference 3. The properties of the array of wires in free space have already been investigated, and the only modification required from the results already obtained is to replace the free-space wavelength of operation, λ_0, by the wavelength, λ, of a plane wave in the solid dielectric. These wavelengths are related by

$$\lambda = \lambda_0/\sqrt{\epsilon_r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The reflection coefficient is required as a function of the angle of incidence i of a plane wave at the surface of the material (Fig. 1), the wave being assumed to have its electric field parallel to the axes of the wires. An equivalent transmission-line circuit for this situation is shown in Fig. 2, and when this is analysed, it is found that the component of the phase-change coefficient in the material parallel to the z-axis, β_z, is given by the equation

$$\cos(\beta_z b) = \cos(\beta b \cos i') + \frac{\pi \sin(\beta b \cos i')}{(\beta b \cos i') \log_e (b/2\pi r)} \quad (4)$$

where

$$\beta = 2\pi/\lambda \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and i' is the inclination to the z-axis of the direction of the plane wave which would propagate in the dielectric if the wires were absent. The angles, i and i', are related by Snell's law

$$\sin i = \sqrt{\epsilon_r} \sin i' \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The effective impedance of the composite material is

$$Z(\beta, i') = \frac{Z_0 \sec i' \tan(\frac{1}{2}\beta b \cos i')}{\sqrt{\epsilon_r} \tan(\frac{1}{2}\beta_z b)} \quad . \quad . \quad . \quad (7)$$

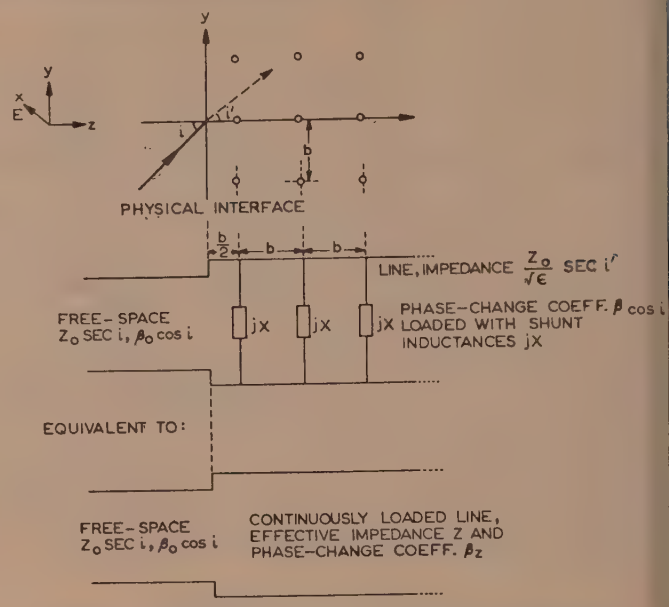


Fig. 2.—Equivalent circuit for free-space/rodded dielectric combination.

The magnetic field of the incident wave lies in the plane of the paper.

and the reflection coefficient at the surface is

$$\rho = \frac{Z(\beta, i') - Z_0 \sec i}{Z(\beta, i') + Z_0 \sec i} \quad . \quad . \quad . \quad . \quad . \quad (8$$

The above equations are derived by the method discussed i Reference 2.

From these equations, the dependence of ρ on the wave length and angle of incidence can be calculated if the parameter of the material ϵ_r, b and r are known. We are, however, con cerned with the synthesis problem of determining the parameter to ensure a small value of ρ over a specified operating range o wavelength and angle of incidence, and this is rather mor difficult. The essential principle of the design is that th dependence on angle of incidence is closely linked to the depen dence on frequency, as may be seen by the appearance of th term $\beta \cos i'$ in the equations. Therefore, if we design th material to have a small reflection coefficient over a wide fre quency range for normal incidence (i.e. $i = i' = 0$), we ma expect to obtain the best achievable performance with respec to angle of incidence. The justification of this procedur follows from an examination of the behaviour, not only for th situation shown in Fig. 1 for which the direction of incidenc is perpendicular to the rods, but also for the case of obliqu incidence when the direction of propagation lies in a plan parallel to the rods and the z-axis. Following establishe waveguide nomenclature, we may refer to these as the H-plan and E-plane cases, respectively. It should be noted that th equations in this Section apply only to the H-plane case.

In the analysis of the following Sections, we first establish procedure for obtaining the best performance over a frequenc band for normal incidence, and then examine the behaviour fo the E-plane and H-plane cases, and thus confirm the validity o the design procedure.

(2.2) Reflection at Normal Incidence

The reflection coefficient for a plane wave incident normall on the surface of the material from free space is obtained b letting i and i' equal zero in eqns. (4), (7) and (8). We a seeking the values of ϵ_r, b and r which will keep the reflectio

oefficient as small as possible over a wide frequency range, and his requires that $Z(\beta, 0)$ should be nearly equal to the free-pace wave impedance, Z_0, over this range. An examination of he equations shows that $Z(\beta, 0)$ varies with β—which is, of ourse, proportional to frequency—in the way shown in Fig. 3(*a*).

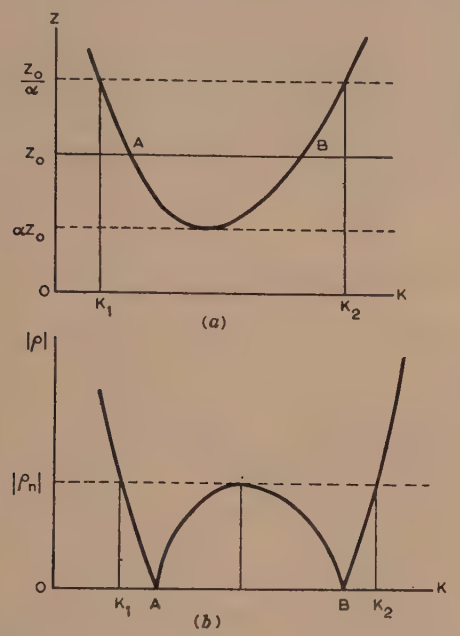


Fig. 3.—Behaviour of rodded dielectric when plane wave is incident normally from free space.

(*a*) Typical dependence on wave impedance, $Z(\beta, 0)$, on β.
(*b*) Corresponding dependence of reflection coefficient on β.

he obvious method of obtaining a small reflection coefficient ver a wide band is to select the minimum value of $Z(\beta, 0)$ to be ss than Z_0 and to correspond to a permitted maximum reflec-on coefficient $|\rho_m|$. This requires

$$Z_{min}(\beta, 0) = \alpha Z_0 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

here

$$\alpha = (1 - |\rho_m|)/(1 + |\rho_m|) \quad . \quad . \quad . \quad . \quad (10)$$

he useful band is then defined by the values β_1, β_2 for which

$$Z(\beta, 0) = Z_0/\alpha \quad . \quad . \quad . \quad . \quad . \quad (11)$$

ne magnitude of the reflection coefficient again being $|\rho_m|$. he phase of the reflection coefficient is π if the frequency lies ithin the range defined by the points A and B for which $|\rho|$ is ero, and it is zero if the frequency is outside this range Fig. 3(*b*)].

The wide-band design is therefore equivalent to finding the alues of ϵ_r, b and r which will give a curve of $Z(\beta, 0)$ against of the required shape. This is most easily done graphically, nd we begin by considering the simpler problem of finding this urve when ϵ_r, b and r are known. Eqns. (4) and (7) are simpli-ed by letting i and i' equal zero.

Define

$$\phi = \beta b \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$A = \pi/\log_\varepsilon (b/2\pi r) \quad . \quad . \quad . \quad . \quad (13)$$

Then, from eqn. (4),

$$\cos(\beta_z b) = \cos\phi + (A/\phi)\sin\phi \quad . \quad . \quad . \quad (14)$$

nd from eqn. (7),

$$Z(\beta, 0) = Z_0 \tan(\phi/2)/\sqrt{\epsilon_r} \tan(\tfrac{1}{2}\beta_z b) \quad . \quad . \quad (15)$$

The term β_z can be eliminated from eqns. (14) and (15) to give $Z(\beta, 0)$ in terms of known quantities.

If

$$\zeta = \sqrt{\epsilon_r} Z(\beta, 0)/Z_0 \quad . \quad . \quad . \quad . \quad . \quad (16)$$

the result of the elimination is

$$\frac{(\zeta^2 - 1)}{A}\phi = \tan(\phi/2) + \frac{\zeta^2}{\tan(\phi/2)} \quad . \quad . \quad . \quad (17)$$

This equation can be used to calculate the values of ϕ and hence of β for which $Z(\beta, 0)$, i.e. ζ, has a specified value. When looked at in this way, the equation is transcendental and can only be solved by numerical methods. A simple approximate graphical solution is, however, possible, since, if ϕ is kept constant, there is a linear relation between $(\zeta^2 - 1)/A$ and ζ^2. This is used to construct the chart shown in Fig. 4, in which we

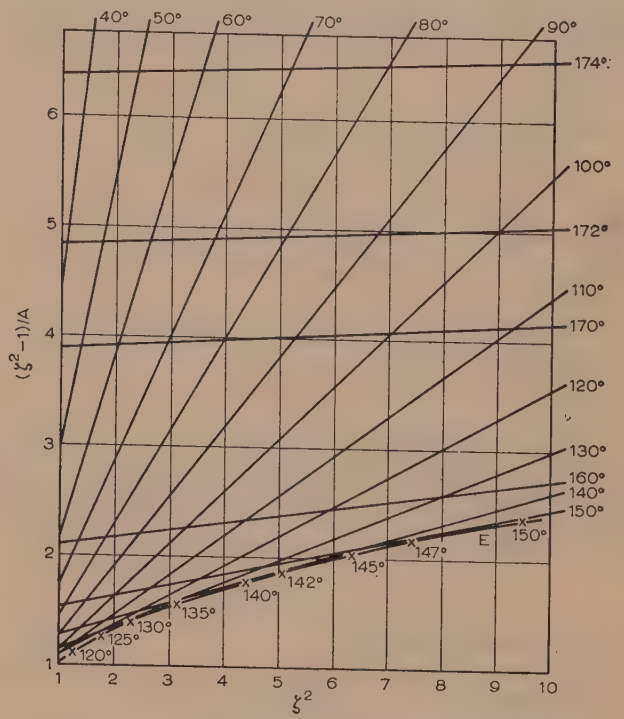


Fig. 4.—Curves to facilitate calculations on rodded media.

The parameter $(\zeta^2 - 1)/A$ is plotted against ζ^2 for different values of ϕ, as shown on each line. E is the envelope of the family of lines.

have a family of straight lines, each defined by a particular value of ϕ. The range of the variables covered by this chart is restricted to those of immediate interest. The curve E is the envelope of the family and has been computed by a standard method.[4] With the aid of this chart we can readily obtain numerical solutions.

(2.2.1) Calculation of the Reflection Coefficient for a Specified Material.

We know the values of ϵ_r, b and r, from which we may calculate A and we wish to plot $|\rho|$ against frequency (or β).

Procedure.

(*a*) Select a value of $|\rho|$.

(*b*) Calculate ζ by using eqns. (8) and (16). Two values of ζ are obtained: for one $Z(\beta, 0) < Z_0$ and for the other, $Z(\beta, 0) > Z_0$.

(*c*) Calculate $(\zeta^2 - 1)/A$ and ζ^2. Plot this point on the chart and read off the values of ϕ for the lines which pass

through the point. Each value of ϕ gives a value of β for which the reflection coefficient has the value selected in (*a*).

In this way, a curve of $|\rho|$ against β can quickly be obtained.

In step (*c*) there are three possibilities, depending on the position of the point relative to the envelope E. If the point is above E two values of ϕ are obtained, and if it is below E there are no values. The situation when the point lies on E is a vital one for the synthesis problem, for now two values of ϕ coalesce corresponding to the minimum value of $Z(\beta, 0)$ in Fig. 3(*a*). To facilitate the determination of ϕ for points on E, a scale is given in Fig. 4 alongside E.

(2.2.2) Design of Material for a Wide-Band Performance.

The available information for the design of material for a wide-band performace will be the maximum tolerable reflection coefficient, $|\rho_m|$, the mid-band wavelength and ϵ_r, the relative permittivity of the supporting structure. The last quantity is selected on considerations of mechanical strength, availability and cost.

Procedure.

(*a*) Calculate α from eqn. (10).

(*b*) Calculate $\zeta_m^2 = \epsilon_r \alpha^2$.

(*c*) From the point on E corresponding to ζ_m^2 read off the values of $\left(\frac{\zeta_m^2 - 1}{A}\right)$ and ϕ. The former leads to A and the latter gives β_m, or the value of β for which $Z(\beta, 0)$ is a minimum.

(*d*) Calculate $\zeta_1^2 = \epsilon_r/\alpha^2$: this corresponds to $Z(\beta, 0) = Z_0/\alpha$. Calculate $(\zeta_1^2 - 1)/A$.

Enter Fig. 4 at $[\zeta_1^2, (\zeta_1^2 - 1)/A]$ and read off the values ϕ_1 and ϕ_2.

Then $\phi_1 = \beta_1 b$ and $\phi_2 = \beta_2 b$, where β_1 and β_2 define the edges of the useful band (see Fig. 3).

(*e*) Let β' be the value of β corresponding to the specified mid-band wavelength. Then, since β is inversely proportional to wavelength,

$$\frac{2}{\beta' b} = \frac{1}{\beta_1 b} + \frac{1}{\beta_2 b} = \frac{1}{\phi_1} + \frac{1}{\phi_2}$$

so that b can be calculated.

(*f*) Use eqn. (13), together with A as determined in (*c*) and b as determined in (*e*), to calculate the rod radius r.

(2.2.3) Numerical Example.

As an example of the procedure in the previous Section, consider

$\epsilon_r = 4$

$|\rho_m| = 0{\cdot}111$

Mid-band free-space wavelength $= \lambda_0$. Then:

(*a*) $\alpha = 0{\cdot}8$.

(*b*) $\zeta_m^2 = 2{\cdot}56$.

(*c*) From Fig. 4, $(\zeta_m^2 - 1)/A = 1{\cdot}46$, $\phi = 132°$. Hence $A = 1{\cdot}07$.

(*d*) $\zeta_1^2 = 6{\cdot}25$, $(\zeta_1^2 - 1)/A = 4{\cdot}91$. From Fig. 4, $\phi_1 = 87{\cdot}9°$, $\phi_2 = 171{\cdot}9°$.

(*e*) $\beta' = 360\sqrt{\epsilon_r}/\lambda_0 = 720/\lambda_0$. Hence $\frac{2\lambda_0}{720b} = \frac{1}{87{\cdot}9} + \frac{1}{171{\cdot}9}$, i.e. $b = 0{\cdot}162\lambda_0$.

(*f*) $A = 1{\cdot}07$ so that $b/2\pi r = \exp(\pi/1{\cdot}07)$, i.e. $r = 0{\cdot}008\,46b$.

This gives all the information required. The variation of the reflection coefficient with operating wavelength can be calculated as in Section 2.2.1 and the results are shown in Fig. 5, the curve having the expected shape.

The useful band is determined by the specified maximum

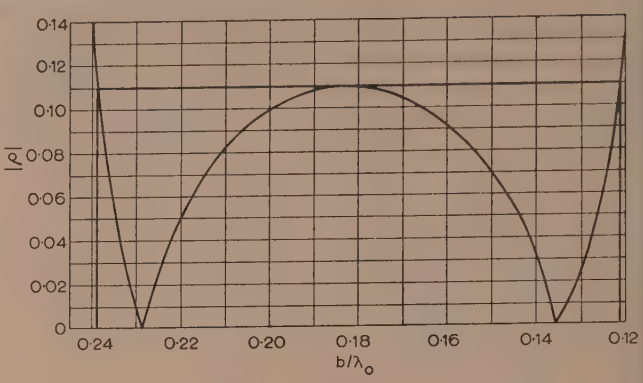


Fig. 5.—Dependence of reflection-coefficient magnitude on b/λ_0 when $b/\lambda_c = 0{\cdot}214$ and $\epsilon_r = 4$.

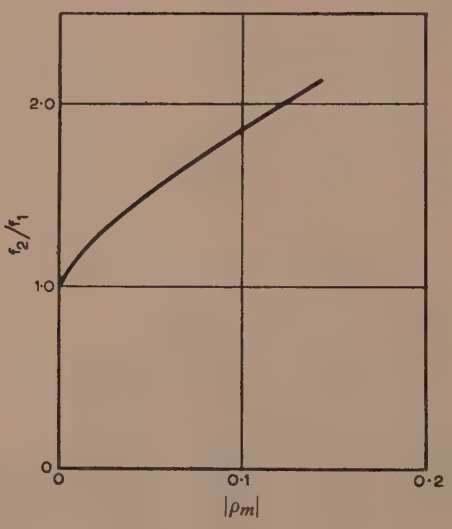


Fig. 6.—Useful bandwidth as a function of maximum permissible reflection coefficient.

$|\rho|$ is less than $|\rho|_m$ for frequencies between f_1 and f_2.

reflection coefficient. A series of calculations has been made, and in Fig. 6 the bandwidth, as indicated by the ratio of the wavelengths at the limits, is plotted against the value of $|\rho_m|$. As would be expected from physical considerations, the bandwidth decreases as $|\rho_m|$ is reduced.

(2.3) Reflection for Oblique Incidence

(2.3.1) *H*-Plane.

The results of the theory in Section 2.1 can be applied immediately to the case of oblique incidence in the *H*-plane. The calculations can be simplified by noting that the wave impedance $Z(\beta, i')$ for incidence at an angle corresponding to i' is simply related to the normal-incidence case for a different wavelength. From eqns. (4) and (7), we have

$$Z(\beta, i') = \sec i' Z(\beta \cos i', 0) \quad . \quad . \quad . \quad (18)$$

The reflection coefficient is determined by the normalized wave impedance, which is

$$Z_H(\beta, i') = Z(\beta, i')/Z_0 \sec i \quad . \quad . \quad . \quad . \quad (19)$$

Eqn. (6) expresses i in terms of i', so that, for normalized impedances, we have

$$Z_H(\beta, i') = \frac{[1 - \epsilon_r \sin^2 i']^{1/2} Z(\beta \cos i', 0)}{Z_0 \cos i'} \quad . \quad . \quad (20)$$

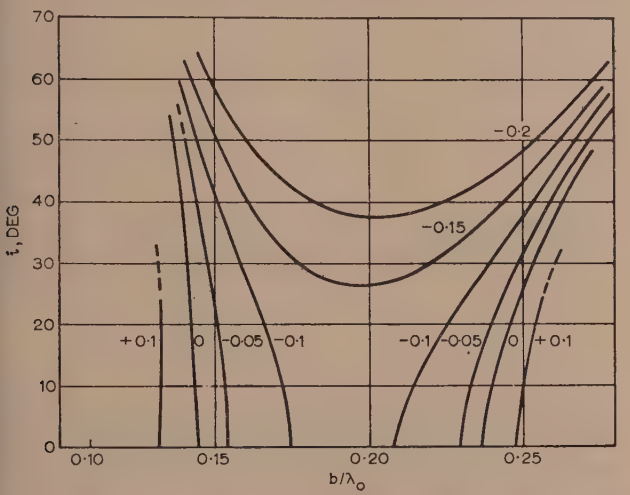
here the suffix H indicates we are considering the H-plane olution.

The reflection coefficient for β, i' is

$$\rho = [Z_H(\beta, i') - 1]/[Z_H(\beta, i') + 1] \quad . \quad . \quad . \quad (21)$$

he curves of Fig. 4 can be used to ease the calculation of eflection coefficients for oblique incidence by noting that ζ, efined in the previous Section, equals $\sqrt{\epsilon_r} Z_H(\beta, 0)$. The intro- uction of ζ as a separate parameter enables the curves of Fig. 4 be drawn in a way which is applicable to all values of ϵ_r.

Contours of constant reflection coefficients are plotted in ig. 7 for the material whose parameters were evaluated in



ig. 7.—Contours of constant reflection coefficient for a free-space-rodded dielectric interface.

H-plane incidence.
$b/\lambda_0 = 0{\cdot}214$.
$\epsilon_r = 4$.

ection 2.2.3. The axes in this diagram are the ratio b/λ_0, λ_0 eing the free-space wavelength, and the angle of incidence i, ese being the quantities of most direct interest.

.3.2) *E*-Plane.

The reflection coefficient for a wave incident in the E-plane an be calculated by using a result derived by Miles.[5] Consider co-ordinate system defined by the x-axis parallel to the rods, e y-axis perpendicular to the rods and lying in the plane of the terface between free space and the material, and the z-axis erpendicular to this interface (see Fig. 1). The interface is the ane $z = 0$. When a plane wave is incident normally to the terface, the solution for the electric field in the material can e written

$$E_x(x, y, z : \beta, 0) = \Theta(y, z, \beta) \exp(j\omega t) \quad . \quad . \quad (22)$$

ere being no variation of field in the x-direction. The extra riables in the functions E_x and Θ are added to facilitate com- rison with the solution for oblique incidence.

Miles has shown that, if the wave in the material is travelling liquely, in such a way that a plane wave in the supporting electric would be travelling in a direction in the x–z plane aking an angle i' with the z-axis, the new solution for E_x is

$$E_x(x, y, z : \beta, i') = \Theta(y, z, \beta \cos i') \exp j(\omega t - \beta x \sin i') \quad (23)$$

e essential point being that the dependence on y and z is the me as for normal incidence if the frequency is changed to cos i'.

The wave impedance of the material at the interface looking in the z-direction is $(E_x/H_y)_{z=0}$, and the corresponding wave impedance for the plane wave incident from free space is $Z_0 \cos i$, i being related to i' by eqn. (6). The normalized wave impedance from which the reflection coefficient can be calculated is therefore

$$Z_E(\beta, i') = \frac{1}{Z_0 \cos i}\left(\frac{E_x}{H_y}\right)_{z=0} \quad . \quad . \quad . \quad (24)$$

From Miles's analysis we find

$$H_y(x, y, z, \beta, i') = \frac{j\sqrt{\epsilon_r}}{\beta Z_0 \cos^2 i'} \frac{\partial}{\partial z} \Theta(y, z, \beta \cos i') \exp(j\omega t - \beta x \sin i') \quad (25)$$

so that

$$Z_E = \frac{j\beta \cos^2 i'}{\sqrt{\epsilon_r} \cos i}\left[\frac{\Theta(y, z, \beta \cos i')}{\frac{\partial}{\partial z}\Theta(y, z, \beta \cos i')}\right]_{z=0} \quad . \quad . \quad (26)$$

This expression is a function of y, but under the conditions in which we are interested, the variation with y is so slight that $Z_E(\beta, i')$ can be regarded as a constant. This is equivalent to assuming that the field in the interface is plane.[2,3] Eqn. (26) is valid for all real angles, including $i' = 0$, so that the normalized impedance for normal incidence is

$$Z(\beta, 0) = \frac{j\beta}{\sqrt{\epsilon_r}}\left[\frac{\Theta(y, z, \beta)}{\frac{\partial}{\partial z}\Theta(y, z, \beta)}\right]_{z=0} \quad . \quad . \quad . \quad (27)$$

By comparing eqns. (26) and (27) and by using eqn. (6) to eliminate i, we obtain

$$Z_E(\beta, i') = \frac{\cos i'}{(1 - \epsilon_r \sin^2 i')^{1/2}} Z(\beta \cos i', 0) \quad . \quad (28)$$

so that, as in the H-plane case, the normalized impedance for oblique incidence at frequency ω is related to the normalized impedance for normal incidence at frequency $\omega \cos i'$. Once again, this enables us to use Fig. 4 to calculate the contours of constant reflection coefficient. They are shown in Fig. 8 for the material of Section 2.2.3.

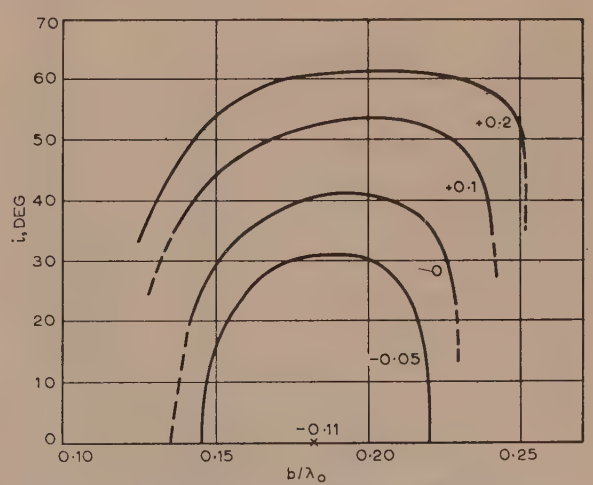


Fig. 8.—Contours of constant reflection coefficient for a free-space/rodded dielectric interface.

E-plane incidence.

(2.3.3) Justification of the Wide-Band Design.

It has been stated that the best overall design is obtained by making the reflection coefficient for normal incidence small over the largest possible range of frequencies. This statement can

be justified by a consideration of eqns. (20) and (28), from which we have

$$Z_H(\beta, i')/Z_E(\beta, i') = (1 - \epsilon_r \sin^2 i')/\cos^2 i' \quad . \quad (29)$$

$$Z_H(\beta, i')Z_E(\beta, i') = Z^2(\beta \cos i', 0) \quad . \quad . \quad . \quad (30)$$

We are seeking a solution in which both Z_E and Z_H are as nearly equal to unity as possible. From eqn. (29), we see that for a given ϵ_r and i the ratio Z_H/Z_E is a constant, C^2, independent of the diameter and spacing of the wires. If $Z(\beta \cos i', 0) = 1$, it is then easily seen that the magnitudes of the reflection coefficients corresponding to Z_H, Z_E are both $|1 - C|/(1 + C)$. However, if $Z(\beta \cos i', 0)$ does not equal unity, it may be shown that at least one of the reflection coefficients exceeds $|1 - C|/(1 + C)$. The best possible solution, i.e. the one for which the reflection coefficient in both planes of incidence is kept small, is that for which $Z(\beta \cos i', 0)$ is as nearly equal to unity as possible. This is the principle which has been used in the design. It may also be noted that eqn. (29) imposes a limit on what can be achieved, and that the form of this equation implies that the performance must deteriorate as i' increases.

(2.4) Choice of Slab Thickness

The analysis has so far been concerned with minimizing the reflection at a single interface. In any practical application, a slab of finite thickness is involved and the number of layers of wires can be selected to cancel the reflections from the two interfaces for the high angles of incidence for which the performance deteriorates. An examination of the present design suggests that three layers is a convenient choice, and curves showing the variation of reflection coefficient with angle of incidence for this case are shown in Figs. 9 and 10.

(3) EXPERIMENTAL VERIFICATION

The theory referred to in the paper can also be used to calculate the dimensions of an array of wires which will increase the wave

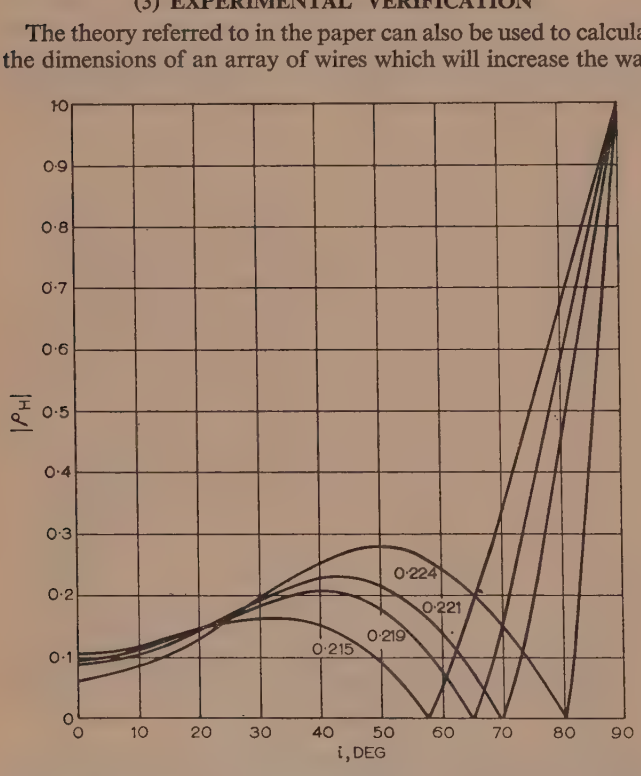


Fig. 9.—Dependence of reflection coefficient on angle of incidence for a slab with three rows of rods.

H-plane incidence.
The parameters on the curves give the values of b/λ_0.

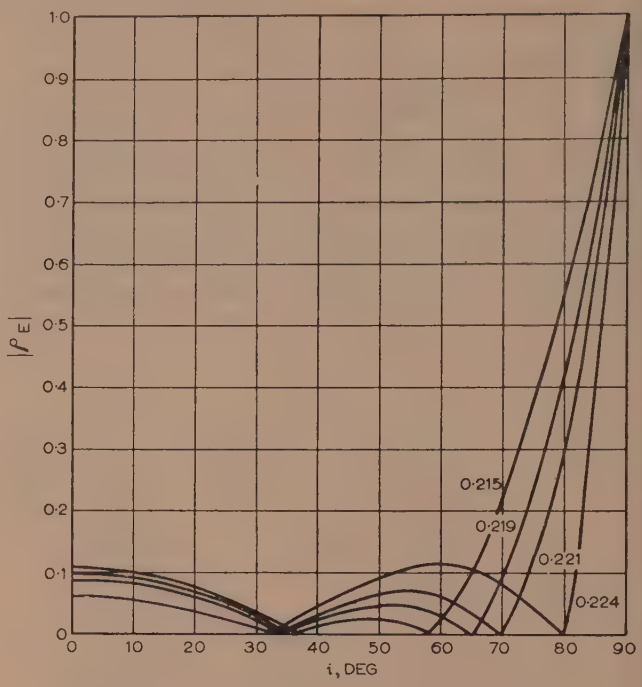


Fig. 10.—Same as Fig. 9, but with *E*-plane incidence.

impedance of a dielectric-filled section of rectangular waveguide to the same value as in an empty section. Calculations were carried out for a polystyrene slab ($\epsilon_r = 2{\cdot}56$) in No. 10 waveguide, the wire spacing b being fixed at 2·40 cm, which is one third of the broad dimension of the guide: the wire diameter, selected to give zero reflection at free-space wavelengths of 8·52 and 12·54 cm, was found to be 0·0144 in, and wire of No. 28 s.w.g. (diameter 0·0148 in) was taken as sufficiently close to this. One row of wires was mounted in a slab of dielectric of thickness b, and the reflection coefficient of the slab was measured over a range of wavelengths, using the Weissfloch method[6, 7] to achieve the necessary accuracy in the small values involved. The experimental values obtained are shown in Fig. 11, compared with the theoretical curve, and it can be seen that the agreement is quite reasonable. The small difference in the values is probably due to the difference between the actual wire diameter and that required theoretically.

The agreement between the measured and theoretical reflection coefficients confirms that the methods of calculation discussed in the paper are adequate for satisfactory design of low-reflection media. The method used for the waveguide case is mathematically identical to that for plane waves incident obliquely in the *H*-plane on an extended grid, so that the results for oblique incidence can also be regarded as being substantiated.

(4) EXTENSION OF THE MEDIUM FOR OPERATION WITH ANY POLARIZATION

The calculations have been based on the assumption that an incident wave is so polarized that the electric field is directed parallel to the wires in the normal-incidence case, and lies in the plane defined by the directions of the axes of the wires and of propagation in the general case of oblique incidence. If the polarization is rotated to make the electric field perpendicular to the wires, the medium behaves as if it were an ordinary dielectric of relative permittivity ϵ_r, the wires having no appreciable effect on the wave propagation. The reflection coefficient for waves of this perpendicular polarization can be reduced by

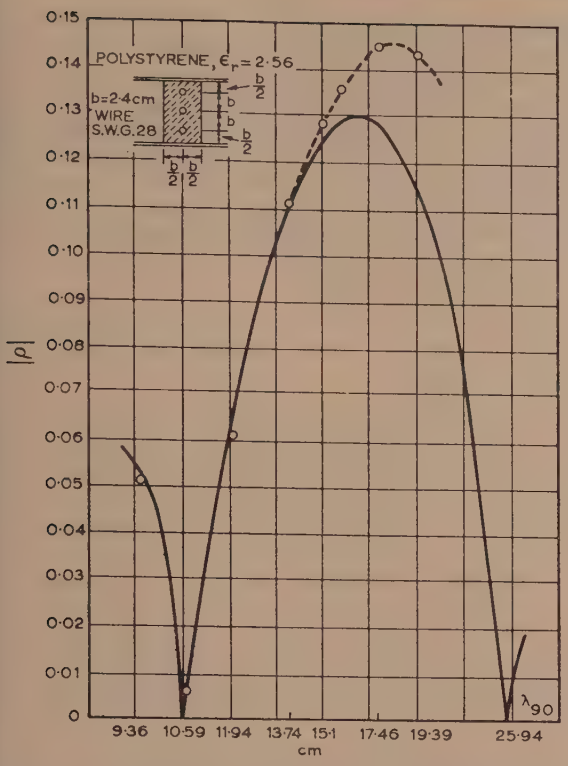


Fig. 11.—Comparison of experimental and theoretical reflection coefficients for waveguide test sample.

—— Theoretical curve.
---- Experimental curve.

cluding a second set of wires identical to the first set and ositioned with their axes perpendicular to those of the first set nd lying in the same planes. Since a wave of arbitrary olarization can be resolved into two component waves having e basic directions of polarization, the composite medium is fective for incident waves of any polarization.

(5) POSSIBLE METHODS OF MANUFACTURE AND TOLERANCES

The method used in the experiments referred to was to split e dielectric slab along the plane of the wires and then to osition the wires under tension in grooves machined in one of e dielectric layers. Such a method is, of course, unsuited for uantity production: the feasibility of alternative methods is verned to a large extent by the tolerances required for the ire diameter and spacings. The wire diameter is obviously ot very critical, since commercial wire of a nominal standard ire gauge proved quite satisfactory for the waveguide experient. The tolerance on the spacing cannot readily be investited, as is the case in all problems of this type: the first point observe is that, since a large number of wires is involved in e construction of a reasonable volume of the material, any nall errors in positioning will, to a certain extent, be averaged ut. Secondly, the major effect of wire incorrectly placed is that e field contribution arising from the current flow in the wire ll be wrongly phased. In aerial design, where fields are nilarly constructed from a large number of independent conbutions, it is usual to specify a tolerance of $2\pi/16$ rad on the ase of each of the individual radiators contributing to the tal field. In the present problem, this is equivalent to a lerance of $\lambda/16$ on the position of each wire, λ being the wavegth measured within the solid dielectric supporting the wires; for example, if this dielectric has a relative permittivity of 4, the tolerance is 1 mm at an operating wavelength of 3·2 cm. One aspect of the positioning problem requires a rather more detailed investigation, i.e. the position of the first row of wires relative to the dielectric surface: the theoretical distance is $b/2$, and a simple analysis shows that, if the actual distance is $(b/2 \pm \zeta)$, the reflection coefficient may be increased by the reflection from a slab of the dielectric of thickness ζ and relative permittivity ϵ_r. An approximate value for the reflection coefficient of such a slab is $4\pi r_1\zeta/(1 - r_1^2)\lambda$, r_1 being the reflection coefficient at a single interface between air and the dielectric) provided that ζ/λ is small. For a dielectric of $\epsilon_r = 4$, r_1 is 1/3, so that the reflection coefficient of the small slab is $3\pi\delta/2\lambda$. The effective distance of the first layer depends on the average value of the spacing errors for the wires contained in it and should be considerably less than $\lambda/16$ for the individual errors; for example, if the average is calculated over a length comprising N wires, δ should be approximately $\lambda/16\sqrt{N}$, i.e. about $\lambda/50$ if $N = 10$, which is equivalent to averaging over a length of some 2·5 wavelengths. This gives a reflection for the slab of about 0·1, corresponding to 1% of the power. The conclusion is that, if the medium is made to tolerances of $\lambda/16$ on the spacing, the reflection coefficient is likely to fluctuate by 0·1 from the values given in the curves. The overall performance of the medium is then likely to be worse than that predicted, but, even so, it should still be better than solid dielectric or sandwich-type materials.

In view of this suggested tolerance of $\lambda/16$ it appears feasible to make slabs of the material by incorporating the wires in the glass-fibre weave used as the basis of many present constructional methods for making radomes. It would be necessary to guard against stretching the weave while it was being surrounded by resin, but this could probably be done, at any rate for relatively flat sheets. The more curved portions of a radome would require special treatment, either by arranging a framework in which wires might be tensioned before the resin was poured on, or by contriving to lay the wire on the surface of sheets of glass fibre of the required shape and thickness. It is obvious that progress in this direction will require a detailed study of manufacturing techniques.

(6) GENERAL DISCUSSION

The work carried out demonstrates that the principle of producing a low-reflection material by embedding wires in a solid dielectric is sound, and that low reflection coefficients can be achieved over quite wide ranges of frequency and angles of incidence. The application of the technique to practical problems requires a study of manufacturing methods and of the changes in behaviour when the interface is not a plane. It is reasonable to suppose that, for interfaces whose radii of curvature are large compared with the wavelength, there will be no marked difference.

A point of considerable interest is the detailed relation between the dependence on wavelength and angle of incidence. This has been noted qualitatively in connection with sandwich-type radomes, but no quantitative relations have been discussed previously. It is probable that similar relations do exist and that their use could lead, as it did in the present problem, to a considerable simplification of the design method.

(7) ACKNOWLEDGMENTS

The authors wish to thank Sir Willis Jackson, Director of Research and Education, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish the paper.

They would also like to thank Prof. H. E. M. Barlow of University College, London, for providing the facilities to carry out the work described.

(8) REFERENCES

(1) CADY, W. M., KARELITZ, M. B., and TURNER, L. A.: 'Radar Scanners and Radomes', Radiation Laboratory Series, Vol. 26 (McGraw-Hill, 1948).

(2) BROWN, J.: 'Artificial Dielectrics having Refractive Indices less than Unity', *Proceedings I.E.E.*, Monograph No. 62 R, May, 1953 (**100**, Part IV, p. 51).

(3) BROWN, J., and JACKSON, WILLIS: 'The Properties of Artificial Dielectrics at Centimetre Wavelengths', *ibid.*, Paper No. 1699 R, January, 1955 (**102** B, p. 11).

(4) EDWARDS, J.: 'Differential Calculus' (Macmillan).

(5) MILES, J.: 'The Reduction of Three-Dimensional Electromagnetic Problems', *Journal of Applied Physics*, 1952, **23**, p. 372.

(6) MARCUVITZ, N.: 'The Waveguide Handbook', Radiation Laboratory Series Vol. 10 (McGraw-Hill, 1951).

(7) BARLOW, H. E. M., and CULLEN, A. L.: 'Microwave Measurements' (Constable, 1950).

DISCUSSION ON THE ABOVE THREE PAPERS BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 24TH NOVEMBER, 1958

Mr. Fuge opened the discussion by asking whether the proposed beam-scanning system could cover the same range of angles as a mechanically scanned aerial. It was stated that, in principle, this could be done, but two limitations were the availability of wide-band electronically-tuned oscillators and the size of the prism needed. The prism fits against the aerial aperture, and its size is therefore governed by the aerial size. For this reason the applications envisaged are those in which relatively small aerials are needed, e.g. in aircraft cloud-warning radars. The principal merit of electronic control is the possibility it offers of high-speed scanning; and a relatively small high-speed scan, requiring either a thin prism or a small frequency change, could be combined with a slower mechanical scan over a wider range of angles. Mr. Millington raised the question of how the oscillator would be modulated, and asked whether this method of scan could be used on both the transmitting and receiving aerials of a radar. The system suggested was to apply a sawtooth frequency-modulation to the transmitter and to use a bank of receivers tuned to different frequencies, so that each received radiation only from a small range of directions. Mr. Fuge pointed out that there would be difficulty in relating direction to frequency, and it was suggested that this could be overcome by a suitable monitoring loop. The possibility of changes in radiated power with the direction of the transmitted beam was raised. The only reasons for such changes would be variations in the transmitter output with frequency, variations in the attenuation through the prism with frequency and the effect of alterations in the beam width as the beam direction changed. Such changes could be kept to within quite small limits.

In reply to questions on the construction of the materials, it was stated that no attempt was made to align the slots in the resonant slot material. There is no evidence that the relative positions of the slots in successive strips have any effect on the refractive index, but this point merits further investigation. Mr. Fuge questioned the need for the solid dielectric, since the conductors played the greatest part in determining the properties. It was pointed out that the dielectric was needed primarily as a mechanical support for the conductors in the dispersive materials. For the radome material, however, the dielectric is essential, since the elimination of reflections is due to a balance between the inductive reactance of the rods and the capacitive action of the dielectric. The position with regard to tolerances was stated in the paper.

Mr. Millington raised a number of points relating to the experimental methods used, which are covered in Reference 5 of the paper by Drs. Seeley and Brown. The experiments were made for the wavelength range 1–10 cm, for which it is anticipated that the materials have practical uses. The prisms used for the measurements were designed to fit in a parallel-plate spectrometer, and were therefore much thinner than those needed in applications. This had the obvious advantage of reducing the effort needed for making test samples. Since the properties of the materials are independent of the thickness, the results are directly applicable to practical situations. It was stressed that the object of the work described was to determine the potentialities of the materials, and that no experimental work had been done on full-size prisms or sheets of radome material.

It was suggested that the figures quoted in the papers for attenuation might be optimistic. This is certainly true for the theoretical figures, since the losses in the supporting dielectric are ignored. On the other hand, the experimental figures are worse than those likely to occur in practical applications, since, as demonstrated in Section 2.2 of the paper by Drs. Seeley and Brown, the contact resistances between the rods and the spectrometer plates contributed a significant proportion of the total measured attenuation. Such contact resistances would not arise in practical applications. The methods of manufacture used in the experimental work were those convenient for laboratory use, and it is probable that improved techniques, which could be used in making large quantities of the material, would further reduce the losses.

Mr. Campbell pointed out that, in the radome material, the wave impedance is made equal to the free-space value and that, if the material is regarded as a dielectric, this implies a relative permittivity of unity. The propagation coefficient should therefore equal the free-space value. Since this does not happen, Mr. Campbell asked for a further discussion of the properties of this material. The answer given first made the point that the concepts of permittivity and permeability must be used with caution in discussing artificial dielectrics. A detailed investigation of this is given in Reference 1 of the paper by Drs. Seeley and Brown, and it has been suggested that the natural parameters to use in defining the properties of an artificial dielectric are the optical ones of refractive index and wave impedance. This has been done in the papers under discussion. It is possible to use the values of refractive index and wave impedance to deduce the permittivity and permeability, but these only have the properties associated with ordinary materials if the artificial dielectric considered is essentially non-dispersive. The materials described in the papers are markedly dispersive. As the frequency tends to zero they cease to behave like dielectrics and approximate to conductors. It is therefore dangerous to base arguments on the ideas of permittivity and permeability, and it is desirable to regard the refractive index and wave impedance as independent parameters. There is therefore no reason why a wave impedance equal to the free-space value should be associated with unity refractive index. An alternative explanation, applicable to the radome material, is that the rods in the material cause an increase in stored magnetic energy. It is to be expected, then, that the relative permeability of the material should be different from unity; in other words, the material cannot be treated as a pure dielectric.

Mr. Cornbleet was intrigued by Figs. 9 and 10 of the paper Mr. Carne and Dr. Brown, in that the curves in Fig. 10 had o nulls whereas those in Fig. 9 had only single nulls. This is explained by reference to Figs. 7 and 8 of the paper. These gures have contours of zero reflection corresponding to the ewster angles, which, in this case, are frequency dependent. ie range of wavelengths used in calculating the curves of gs. 9 and 10 is such that the Brewster phenomenon occurs in the latter (the E-plane situation) at angles around 35°, but not in the former (the H-plane situation). It will be noted that, for a given wavelength, the zeros in the vicinity of 70° occur at precisely the same angle in both Figures. These zeros correspond to the angle for which the effective phase shift through the slab is 180°; this, of course, is independent of the plane of incidence. The thickness of the slab was chosen to make the angles at which this 180° phase shift occurs as close to 90° as possible.

621.372

The Institution of Electrical Engineers
Paper No. 2737 E
Mar. 1959

AN INVESTIGATION OF THE EXCITATION OF RADIATION BY SURFACE WAVES

By K. P. SHARMA, M.Sc., Ph.D., Graduate.

(*The paper was first received* 26*th April, and in revised form* 19*th July*, 1958.)

SUMMARY

The graphical method of measuring a reflection coefficient through a junction, originated by Deschamps, is used to determine the transmission efficiency of a surface wave at a discontinuity. The radiations excited by the surface wave at a discontinuity in surface reactance, and at an edge of a metallic strip above a reactive surface, are investigated. It is shown that the discontinuities cause the radiation of an appreciable fraction of the incident power. A matching step at the discontinuity in surface reactance is found to have no effect on the excitation of radiation. Further, it is found that, compared with the radiation excited at the edge of a metallic strip, the radiation excited at a discontinuity in surface reactance is confined to a narrower angle above the surface.

LIST OF PRINCIPAL SYMBOLS

E_x, E_y = Components of the electric field.
H_z, H_ϕ = Components of the magnetic field.
k = Phase-change coefficient of a plane wave in free space.
X_s = Surface reactance of a guiding surface.
β_s = Surface-wave phase-change coefficient.
u = Surface-wave vertical decay coefficient.
μ_0 = Permeability of free space.
ϵ_0 = Permittivity of free space.
$H_n(\ldots)$ = Hankel function.
ω = Angular frequency.

(1) INTRODUCTION

Surface waves[1] are electromagnetic waves which propagate without radiation along the interface between two different media. The problem of launching these waves has received considerable attention. The investigation of the excitation of radiation from these waves has so far been confined mainly to the case in which the guiding structure is terminated either by a smooth metallic surface[2, 3] or by free space.[4] Simon and Weill[5] and Mueller[6] have described aerials that use the radiation excited by surface waves at a series of discontinuities to build up an end-fire or a broadside array. A general study of the excitation of radiation by a surface wave at a discontinuity in its path of propagation has not so far been made. The object of the paper is to give an account of such an investigation.

The excitation of radiation by a surface wave at a discontinuity in surface reactance, and that at an edge of a metallic strip above a reactive surface, have been studied particularly. The graphical method of measuring a reflection coefficient through a junction, originated by Deschamps,[7] has been applied to determine the transmission efficiency of the two types of discontinuities. The polar distributions of the fields above the surface have been obtained with and without the discontinuity in order to obtain information about the scatter of the radiation excited by the surface wave.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

(2) THEORETICAL CONSIDERATIONS

(2.1) Basic Field Equations

The present investigation is concerned with surface waves launched over a flat reactive surface. The two forms of surface waves that are known to be supported by a flat reactive surface are the inhomogeneous plane surface waves, and the radial cylindrical surface waves (Fig. 1). The field components of these waves are as follows:

For Plane Surface Wave

$$H_z = A \exp(-uy - j\beta_s x) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$E_x = Z_s A \exp(-uy - j\beta_s x) \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$E_y = \frac{\beta_s}{k}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)} A \exp(-uy - j\beta_s x) \quad . \quad . \quad (3)$$

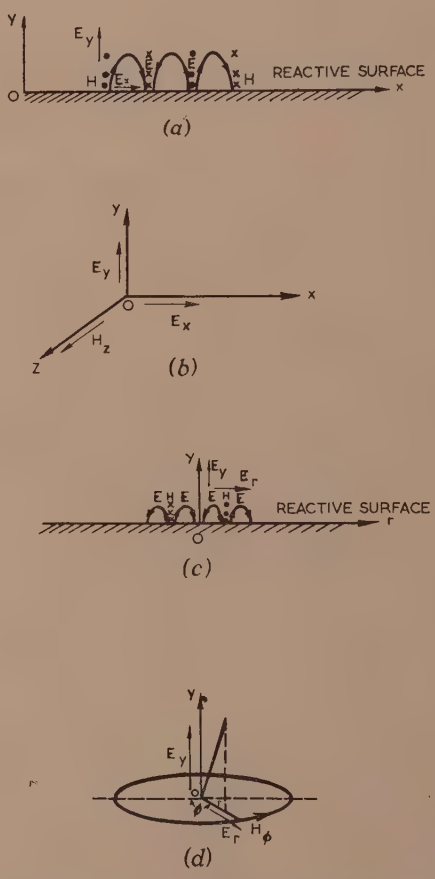


Fig. 1.—Surface waves on a flat surface.
(*a*) Plane surface wave.
(*b*) Cartesian co-ordinates, showing the field components.
(*c*) Radial cylindrical surface wave.
(*d*) Cylindrical co-ordinates, showing the field components.

For Radial Cylindrical Surface Wave

$$H_\phi = A \exp(-uy) H_1^{(2)}(\beta_s r) \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$E_r = Z_s A \exp(-uy) H_1^{(2)}(\beta_s r) \quad . \quad . \quad . \quad . \quad (5)$$

$$E_y = \frac{\beta_s}{k}\sqrt{\left(\frac{\mu_0}{\epsilon_0}\right)} \exp(-uy) H_0^{(2)}(\beta_s r) \quad . \quad . \quad . \quad (6)$$

where $$\beta_s^2 = k^2 + u^2 = \omega^2\mu_0\epsilon_0 + u^2 \quad . \quad . \quad . \quad (7)$$

and $$Z_s = \left(\frac{E_x}{H_z}\right)_{\gamma=0} = -\left(\frac{E_r}{H_\phi}\right)_{\gamma=0} = \frac{ju}{\omega\epsilon_0} = jX_s \quad . \quad (8)$$

The factor $\exp j\omega t$ has been suppressed in the field equations.

(2.2) Surface-Wave Transmission Efficiency at a Discontinuity

(2.2.1) Discontinuity in Surface Reactance.

When X_{s1} is less than X_{s2} (Fig. 2), the expression for the fraction of the incident power which is transmitted, obtained by Brown,[8] is

$$T = \frac{4u_1u_2}{(u_1+u_2)^2}\,\frac{\beta_{s1}+u_1}{\beta_{s2}+u_2} \quad . \quad . \quad . \quad . \quad (9)$$

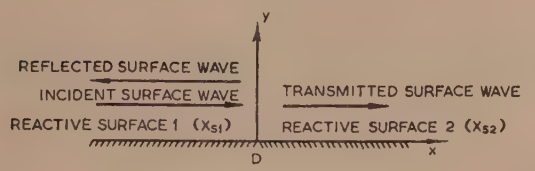


Fig. 2.—Discontinuity in surface reactance.
The surface reactance of the plane $y = 0$ has a discontinuity at D.

The fraction of the incident power which is reflected is given by either

$$R_1 = \frac{u_1^2(u_2-u_1)^2}{\beta_{s1}^2(\beta_{s2}+\beta_{s1})^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

or $$R_2 = \frac{u_2^2(u_2-u_1)^2}{\beta_{s2}^2(\beta_{s2}+\beta_{s1})^2}\left(\frac{\beta_{s1}+u_1}{\beta_{s2}+u_2}\right)^2 \quad . \quad . \quad . \quad (11)$$

The suffixes 1 and 2 in R imply that the wave is incident from the side 1 or 2, respectively.

When X_{s1} is greater than X_{s2} the factor $(\beta_{s1}+u_1)/(\beta_{s2}+u_2)$ is inverted in the expression for T and it disappears from R_2, appearing again inverted in R_1.

The results of numerical evaluation of R_1, R_2 and T are shown in Fig. 3 for the case when $X_{s1} = 58$ ohms. It will be observed that the coefficients R_1 and R_2 are so small that T is approximately equal to $\eta(t)$, the transmission efficiency, defined as the ratio of the output power to the input power, irrespective of the side from which the wave is incident.

A detailed theoretical analysis of the related problem of diffraction at the edge of a dielectric slab has been given by Angulo.[9]

(2.2.2) Metallic Strip above a Reactive Surface.

An approximate expression for the surface-wave transmission efficiency of a metallic strip above a reactive surface has been obtained on the assumption of a surface-wave type of aperture distribution above the edge of the metallic strip (Fig. 4). This approach is similar to one used by Cullen[10] to calculate the launching efficiency of a dropped surface-wave source. On this basis, the transmitted surface-wave field is given by

$$H_z^s = \frac{k}{\beta_s}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\mathscr{E}_H \exp[-u(y+H) - j\beta_s x] \quad (12)$$

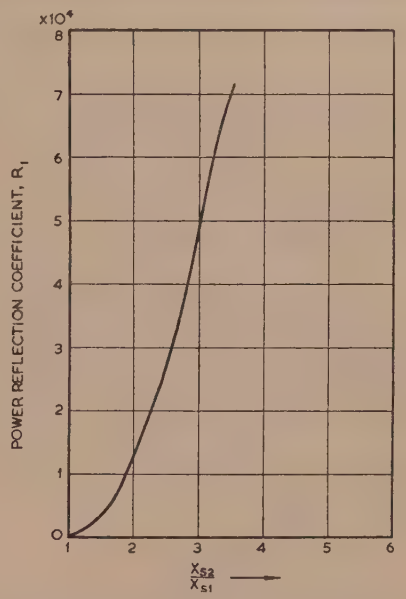


Fig. 3.—Dependence of R_1, R_2 and $\eta(t)$ on the ratio of surface reactances X_{s2}/X_{s1}.
$X_{s1} = 58$ ohms. Frequency: 9·425 Gc/s.

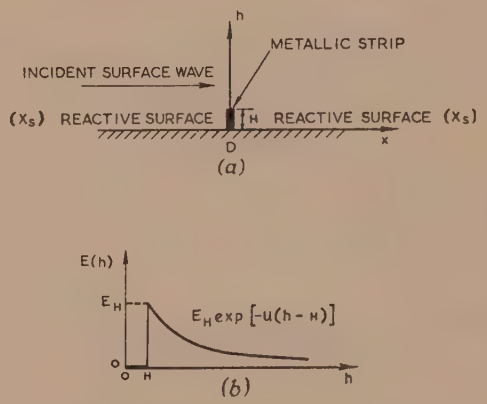


Fig. 4.—Metallic strip above a reactive surface.
(*a*) The strip and the co-ordinates.
(*b*) Aperture distribution in and above the plane of the strip.

and the power in the surface per unit wavefront is given by

$$P_s = \frac{1}{4}\,\frac{k}{\beta_s u}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)} \exp(-2uH)|\mathscr{E}_H|^2 \quad . \quad . \quad (13)$$

where H is the height of the edge above the surface, and $\mathscr{E}_H$ is defined by the field distribution on the plane surface of the discontinuity by

$$\mathscr{E}(h) = \mathscr{E}_H \exp[-u(h-H)] \text{ at } h \geqslant H \quad . \quad . \quad (14)$$

The far-zone radiation field is given by

$$H_z^R = 2jk\sqrt{\left(\frac{2}{\pi}\right)}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}\mathscr{E}_H \frac{\exp(-jkr - \pi/4)}{(kr)^{1/2}} F(\theta) \quad (15)$$

where $$F(\theta) = \frac{\sin(kH\sin\theta)}{u - jk\sin\theta} \quad . \quad . \quad . \quad . \quad (16)$$

so that $$|F(\theta)|^2 = \frac{\sin^2(kH\sin\theta)}{u^2 + k^2\sin^2\theta} \quad . \quad . \quad . \quad . \quad (17)$$

The power carried by the radiated wave, per unit wavefront, is given by

$$P_R = \frac{k}{\pi}\sqrt{\left(\frac{\epsilon_0}{\mu_0}\right)}|\mathscr{E}_H|^2 \mathrm{I}(\theta) \quad . \quad . \quad . \quad . \quad (18)$$

Here,

$$\mathrm{I}(\theta) = \int_0^{\pi/2} |\mathrm{F}(\theta)|^2 d\theta \quad . \quad . \quad . \quad . \quad . \quad (19)$$

By definition, the transmission efficiency is

$$\eta(t) = \frac{P_s}{(P_s + P_R)} = \frac{\pi \exp(-2uH)}{\pi \exp(-2uH) + 4\beta_s u \mathrm{I}(\theta)} \quad . \quad (20)$$

Numerical values of $\eta(t)$ have been calculated for different values of H for the case when $u = 0{\cdot}303$/cm, and $\beta_s = 1{\cdot}997$ rad/cm, the results being shown in Fig. 5.

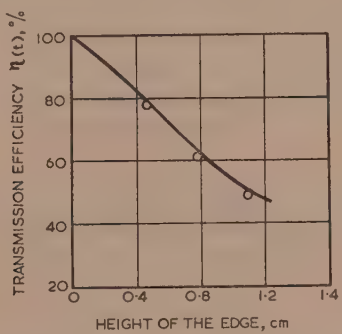


Fig. 5.—Dependence of $\eta(t)$ on the height of the edge of the strip over the reactive surface.
$X_s = 58$ ohms.
$f = 9{\cdot}425$ Gc/s.
○ = Experimental points.

The integral $\mathrm{I}(\theta)$ was evaluated by numerical method, but in a recent paper Cullen[10] evaluated the same integral using an infinite series.

(3) EXPERIMENTAL DETERMINATION OF SURFACE-WAVE TRANSMISSION EFFICIENCY

(3.1) Outline of the Principle

A discontinuity, of the type under consideration, in the path of propagation of a surface wave may be looked upon as a four-terminal passive and linear network connection (Fig. 6) between

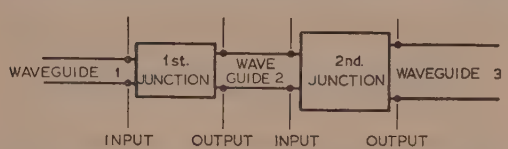


Fig. 6.—Network connection representing two discontinuities.
Junction 1.—Contains a coaxial-to-American-rectangular-waveguide transformer, a piece of coaxial line, and a half-wave dipole.
Junction 2.—Contains a discontinuity in the path of a surface wave.
Waveguide 1.—American-rectangular waveguide.
Waveguide 2.—Surface waveguide of reactance X_{s1}.
Waveguide 3.—Surface waveguide of reactance either X_{s1} or X_{s2}.

two waveguides. In terms of the scattering coefficients of the junction, the transmission efficiency is given by either

$$\eta_1(t) = \frac{\text{Output power}}{\text{Input power}} = \frac{|S_{12}|^2}{1 - |S_{11}|^2} \quad . \quad . \quad (21)$$

or

$$\eta_2(t) = \frac{|S_{12}|^2}{1 - |S_{22}|^2} \quad . \quad . \quad (22)$$

the suffixes 1 and 2 implying that the wave is incident from the side 1 or 2, respectively.

The technique which has been applied in the present investigation is the graphical technique as presented by Storer, Sheingold and Stein.[11] Du Hamel[13] suggested a similar approach in determining the launching efficiency of the plane surface-wave launcher of Rich,[12] who has quoted some results based on this method.[13]

In the present investigation, the scattering coefficients of the junction between a rectangular guide feeding the surface-wave launching aerial and the corrugated-surface waveguide is first determined. The discontinuity is then introduced in the path of the surface wave, and the reflection coefficients measured in the feed guide are analysed, in the way suggested by Storer, Sheingold and Stein,[11] with a view to transferring them to an output reference plane situated above the corrugated surface. These are then transferred near the discontinuity.

When the wave transmitted across a discontinuity in its path of propagation is terminated by a short-circuit, the reflection coefficient at the output reference plane of the junction (Fig. 6) containing the discontinuity is given by

$$\rho = \exp j(\phi + 2\beta_s l) \quad . \quad . \quad . \quad . \quad . \quad (23)$$

This equation plots itself into a circle of unit radius in the complex ρ-plane [Fig. 7(*a*)] as $\beta_s l$ is varied. The transformation

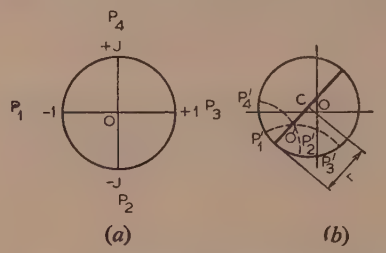


Fig. 7.—Reflection coefficient circles.
(*a*) Unit. (*b*) Image.

of ρ at the output reference plane to ρ' at the input reference is given by the equation

$$\rho' = \frac{S_{11} + (S_{12}^2 + S_{11}S_{22})\rho}{1 - S_{22}\rho} \quad . \quad . \quad . \quad (24)$$

where S_{11}, S_{12} and S_{22} are the scattering coefficients of the second junction. This equation represents a bilinear transformation. Hence the image of the unit ρ-circle is a ρ'-circle of radius r (Fig. 7*b*). It has been shown[10] that r is given by

$$r = \frac{|S_{12}|^2}{1 - |S_{22}|^2} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

Therefore $\eta_2(t) = r$.

For a symmetrical junction $|S_{11}| = |S_{22}|$ so that $\eta_1(t) = \eta_2(t)$, and r gives the transmission efficiency irrespective of the side from which the wave is incident. For an asymmetrical junction $|S_{11}|$, $|S_{12}|$ and $|S_{22}|$ must be determined separately to calculate $\eta_1(t)$ and $\eta_2(t)$.

(3.2) Discussion of the Technique of Measurement

The application of the principle, outlined in the previous Section, to the problem of determining the surface-wave transmission efficiency, is justified only if the launching aerial receives a pure surface wave, the radiated waves that originate from the launching aerial and from the discontinuity in the path of the surface wave being lost in space. In an experimental arrangement it will be necessary to use a launcher of high efficiency.

The choice of a radial cylindrical surface wave, which has been found to be launched with an efficiency of 80%, seems to be obvious on this ground. The radiation fields picked up by the aerial can be minimized by having a large separation between the discontinuity and the launching aerial, and also a large separation between the discontinuity and the short-circuit termination to the waves transmitted across the discontinuity. The justification for this arrangement has been provided by a series of standing-wave patterns obtained with and without the discontinuity, showing that the contamination of the surface wave at the surface by the radiation field falls off gradually with distance measured from the source of the radiation field. Thus, if the launching aerial is situated as near the surface as the condition of high launching efficiency may permit, this will receive very little power from the radiated wave, so that the experimental arrangement will very nearly fulfil the requirement that the radiated waves are lost in space.

It is obvious that the suggested system necessitates the use of a fixed short-circuit termination. Therefore the electrical distance between the short-circuit and a fixed reference plane has to be changed by frequency variation. For the application of the method of determining transmission efficiency, outlined in Section 3.1, it is necessary for the first junction as well as the second, which contains the discontinuity, to be frequency independent over a narrow band of frequency. In practice, it may be possible sometimes to make the first junction frequency independent over a narrow frequency band, and this narrow band may be sufficient to provide the necessary data. Thus, provided that the discontinuity is not too frequency sensitive, the suggested system should be adaptable for the purpose of determining the transmission efficiency of a surface wave at the discontinuity.

(3.3) Experimental Arrangement

The experimental arrangement used for the transmission-efficiency determinations of the two types of the discontinuities studied are shown in Fig. 8. The junction between the feed

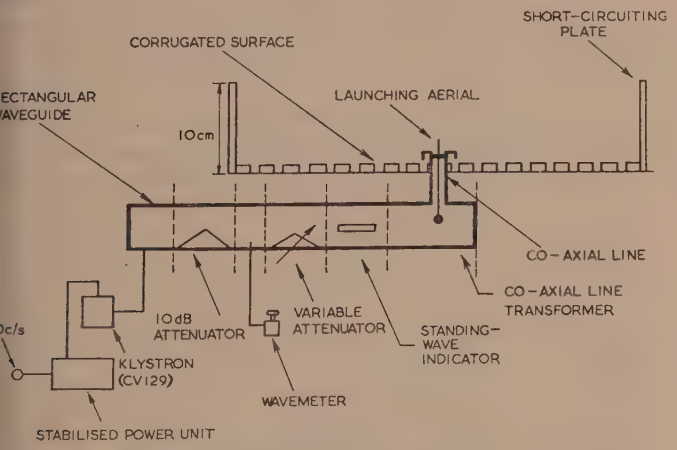


Fig. 8.—Experimental arrangement.

guide and the corrugated surface waveguide, i.e. the first junction, contains a rectangular-waveguide-to-coaxial-line transformer, a short piece of coaxial line and the half-wave dipole launcher. The rectangular-waveguide-to-coaxial-line transformer provides a matched junction between the American guide and the 72-ohm-impedance coaxial line. The distance between the short-circuit termination to the surface and the launching aerial is about 72 cm. The standing-wave pattern plotted over a distance of 15 wavelengths from the short-circuit to the surface wave shows that the minima of the pattern are never distinguishable from the zero signal level. Hence the resistive attenuation of the guiding system is negligible.

A discontinuity in surface reactance was introduced by coating the corrugated surface, over a distance of about 31 cm from the short-circuit termination, with thin sheets of polystyrene. The metallic strip discontinuity above the corrugated surface is obtained by holding a 0·005 in-thick brass strip against the wall of a corrugation by means of a few small polystyrene packing pieces so that it stands vertically over the surface and normal to the incident wave.

The short-circuit termination is provided by an aluminium cylinder, of height 4 in, over the corrugated metallic disc. This has been located in one of the corrugations and held tightly against the wall and the bottom of the corrugation by means of square wooden blocks clamped to the corrugated metallic disc. The height of the reflector above the surface is sufficient to cover almost the entire spread of the surface wave.

A resistive termination to the surface wave, whenever necessary, was achieved by a load made from strawboard coated with Aquadag.

(3.4) Experimental Results

(3.4.1) Scattering Coefficients of the First Junction.

The scattering coefficient S_{11} is determined by terminating the corrugated structure by a resistor and measuring the reflection coefficients at the input reference plane over a frequency band of ± 110 Mc/s about 9·425 Gc/s. Fig. 9 shows these data. It is clear that the resistive termination has a small reflection coefficient which combines with the small reflection coefficient of the junction so that the reflection-coefficient data obtained over the narrow frequency band are distributed about the circumference of a circle. The reflection coefficient of the junction, i.e. S_{11}, is easily deducible from Fig. 9.

The reflection-coefficient circle obtained at the input reference plane by terminating the surface wave with a short-circuit is shown in Fig. 10. The datum points are well distributed about the circumference of a circle, so that the junction between the rectangular waveguide and the corrugated surface waveguide is very nearly frequency independent over the narrow frequency band of ± 110 Mc/s about 9·425 Gc/s. The output reference plane (Fig. 6) has been chosen to correspond to an electrical short-circuit at a frequency of 9·425 Gc/s. The graphical analysis[10] has been used to obtain the following values of the coefficients S_{11}, S_{12} and S_{22}:

$$S_{11} = 0{\cdot}049 \angle 141°0$$
$$S_{12} = 0{\cdot}843 \angle 88°5'$$
$$S_{22} = 0{\cdot}121 \angle 162°0$$

Assuming that the power loss from the junction is entirely due to radiation, the surface-wave launching efficiency, η, of the aerial is given by

$$\eta = \frac{|S_{12}|^2}{1 - |S_{11}|^2} = 0{\cdot}712 = 71{\cdot}2\%$$

This value differs very little from the value of 72·8% obtained by the method of Fernando and Barlow.[13]

(3.4.2) Transmission Efficiency of a Discontinuity in Surface Reactance

Fig. 11 shows the reflection-coefficient data obtained by using a short-circuit termination for the case when a discontinuity in the surface reactance was achieved by coating the corrugated surface with 0·050 in-thick polystyrene sheets beyond a distance of 40·97 cm from the centre. The power reflection coefficient, R_1, of the discontinuity is so small that $\eta_1(t)$ is very nearly equal to the power transmission coefficient of the discontinuity. Analysis of the data also shows that the power

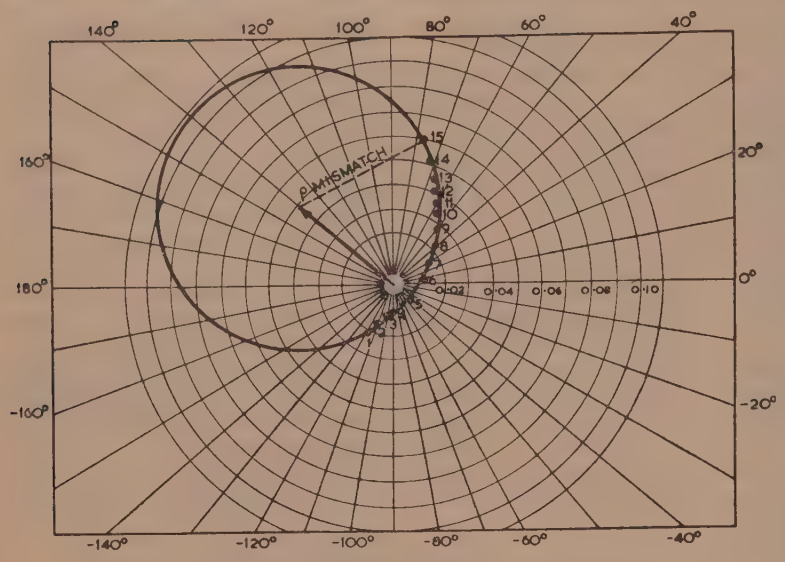


Fig. 9.—Reflection-coefficient circle.
Resistive termination to the surface wave ($X_s = 58$ ohms).

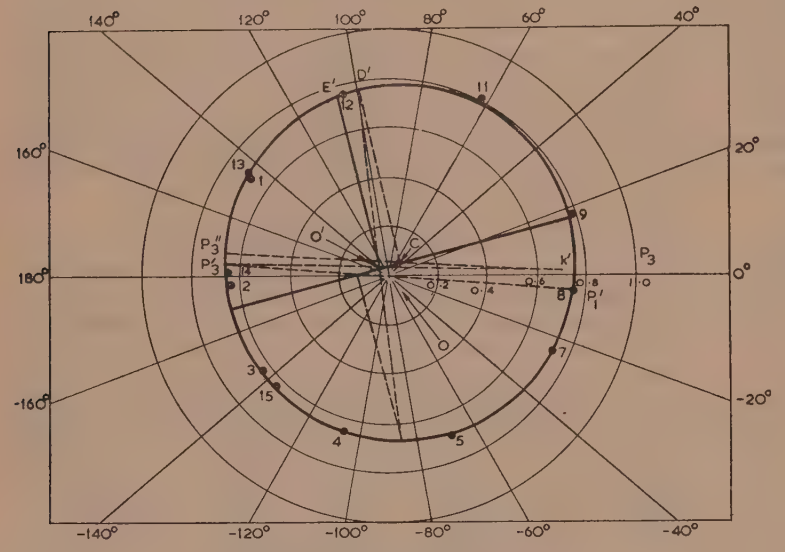


Fig. 10.—Reflection-coefficient circle.
Short-circuit termination to the surface wave ($X_s = 58$ ohms).

reflection coefficient R_2 is bigger than R_1. Nevertheless, it is not large enough to affect the transmission efficiency of the discontinuity significantly. Thus the transmission efficiency of the discontinuity is approximately given by the radius of the reflection-coefficient circle in Fig. 11. These experimental results confirm the theoretical results mentioned in Section 2.2.1.

The reactance of the surface obtained by coating the 58-ohm corrugated surface with 0·050 in-thick polystyrene sheets is 129·7 ohms. Hence the magnitude of the discontinuity in the surface reactance is

$$\frac{X_{s2}}{X_{s1}} = 2{\cdot}24$$

The transmission efficiencies of a few more similar discontinuities having different values of X_{s2}/X_{s1} have been determined, the result being shown in Fig. 3. It will be observed that all the experimental points, except one corresponding to $X_{s2}/X_{s1} = 2{\cdot}49$, show good agreement with the theoretical results. Further, it will be seen from Fig. 12 that a few of the reflection-coefficient datum points in this case fall below the average circle drawn through most of them. A similar effect, although not very marked, can also be seen in Fig. 11. Measurements repeated by changing the position of the discontinuity have shown that this simply rotates the Figure by an appropriate amount. This confirms that the radiation excited from the reflected surface wave is not picked up by the aerial. Further measurements over a wider frequency band show that the distortion of the circle is repeated at a frequency interval of about 180 Mc/s. This frequency interval indicates the presence of a secondary effect which involves a path length of about 167 cm, which is close to the path traversed by a wave from the discontinuity to the launching aerial, via the reflecting termination, and from the aerial to the discontinuity. This suggests that a fraction of the power radiated from the discontinuity was being picked up by the aerial after being reflected from the short-circuit termination, and then relaunched by the aerial as a surface wave. It is difficult to assess the effect of this on the technique of determining the transmission efficiency of a surface wave at a discontinuity in surface reactance. For the case in which the discontinuity is given by $X_{s2}/X_{s1} = 2{\cdot}49$, the radiated wave picked up by the launching aerial may account for the discrepancy between the experimental and the theoretical results. However, the agreement between these results for the reactance ratio $X_{s2}/X_{s1} = 2{\cdot}24$ indicates that the technique is applicable for discontinuities for which $X_{s2}/X_{s1} < 2{\cdot}3$. The geometrical discontinuity between the two reactive surfaces, producing a discontinuity in surface reactance given by $X_{s2}/X_{s1} = 2{\cdot}24$, is 1·3 mm, whereas the same for a discontinuity in surface reactance given by $X_{s2}/X_{s1} = 2{\cdot}49$ is 1·5 mm. Therefore, it seems unlikely that the geometrical discontinuity is responsible for the extra radiation loss for the case in which the discontinuity in surface reactance is given by $X_{s2}/X_{s1} = 2{\cdot}49$.

(3.4.3) Effect of a Step at a Discontinuity in Surface Reactance.

A dielectric (polystyrene) step, a quarter of a free-space wavelength long and of half the thickness of the

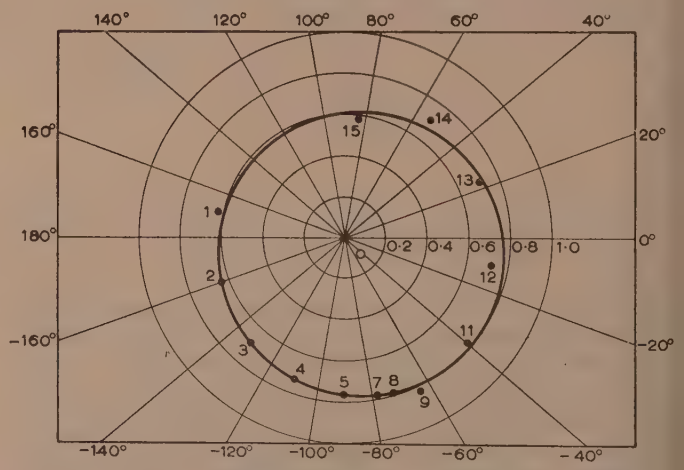


Fig. 11.—Reflection-coefficient circle at the discontinuity in surface reactance.
$X_{s2}/X_{s1} = 2{\cdot}24$ with a short-circuit termination to the surface wave. Discontinuity at 40·97 cm.

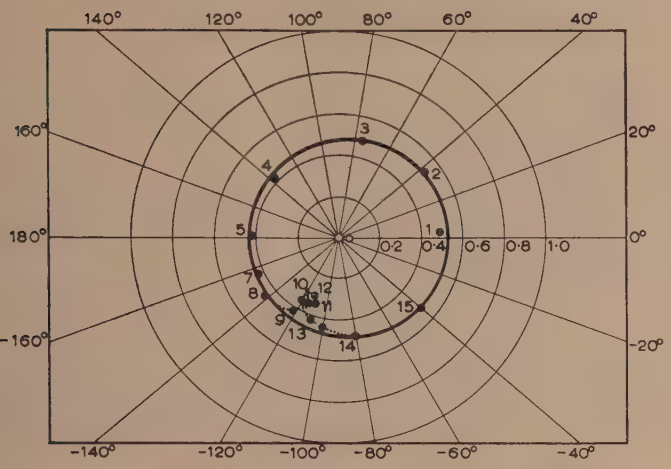


Fig. 12.—Reflection-coefficient circle at the discontinuity in surface reactance.
$X_{s2}/X_{s1} = 2{\cdot}49$ with a short-circuit termination to the surface wave.
Discontinuity at 40·97 cm.

dielectric coating over the corrugated surface, was arranged before the discontinuity in the surface reactance. The transmitted surface wave was terminated by a short-circuit. The radius of the reflection-coefficient circle obtained with this arrangement is the same as that of the circle obtained without introducing the step. The step has, therefore, no effect on the excitation of radiation by the surface wave. This seems to suggest that a small taper at the discontinuity will not reduce the radiation loss.

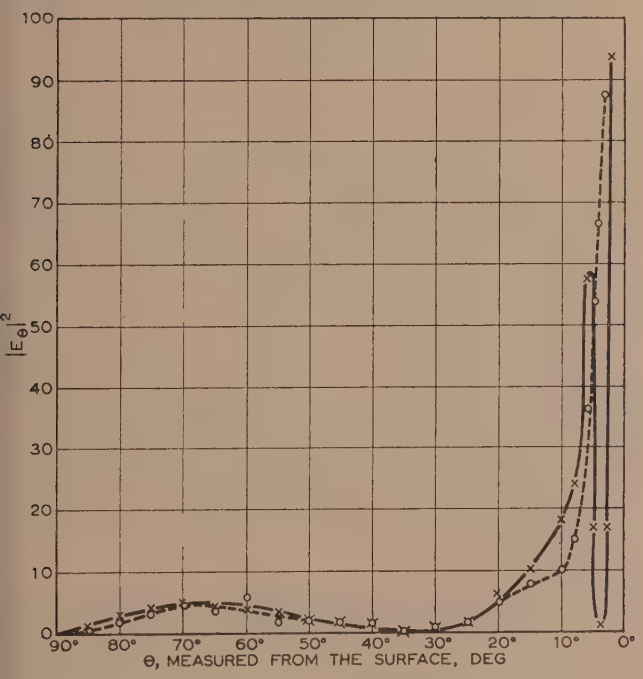


Fig. 13.—Polar power-flow density pattern with and without a discontinuity in surface reactance.
$X_{s2}/X_{s1} = 2{\cdot}24$.
× — × With discontinuity at 40·97 cm.
○ · · · ○ No discontinuity.
$X_{s2}/X_{s1} = 2{\cdot}24$.
Radial distance of the measuring probe = 56 cm.

(3.4.4) Transmission Efficiency of a Metallic Strip above a Reactive Surface.

The reflection-coefficient circles obtained from the measurements on metallic strips (with short-circuit termination to the surface waves) are similar to that in Fig. 11. The radii of the circles give the transmission efficiencies directly. As the width of the strip increases, more and more points fall away from the mean circle, showing that the discontinuity becomes more and more frequency sensitive. The transmission efficiencies of metallic strips of three different widths are shown in Fig. 5. The type of radiation pick-up effect observed when there is a discontinuity in surface reactance has not been encountered in this case. This may be because the radiation excited is not so strongly confined to the surface.

(4) POLAR DISTRIBUTION OF THE RADIATION FIELD EXCITED FROM A SURFACE WAVE

The pattern of the radiation excited from a surface wave cannot be obtained directly by an experimental procedure. Nevertheless, valuable information can be extracted from the comparison of the polar-distribution measurements of the electromagnetic fields with and without a discontinuity. It will be observed from Figs. 13 and 14 that the radiation field excited

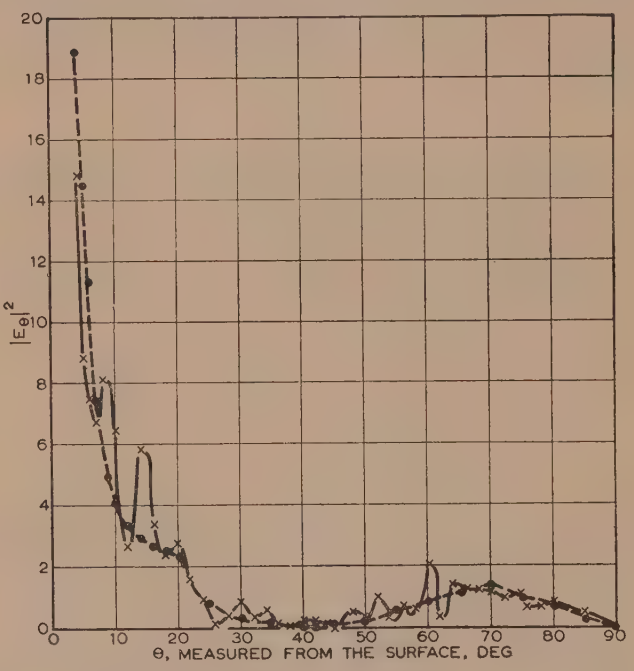


Fig. 14.—Polar power-flow density pattern with and without a metallic strip (0·47 cm wide) above a reactive surface ($X_s = 58$ ohms).
× — × With edge at 41·25 cm.
● — ● No discontinuity.
Height of the edge = 0·47 cm.
Radial distance of the measuring probe = 53 cm.

by the surface wave is more confined near the surface when there is a discontinuity in surface reactance than when there is a metallic strip above the reactive surface.

(5) CONCLUSIONS

The experiments described demonstrate the possibility of determining the transmission efficiency of a discontinuity in the path of a surface wave, and confirm the theoretical value.[8] The theoretical transmission efficiency of a metallic strip over a

reactive surface has been found to agree, within the limits of experimental error, with the experimental result. A matching step at a discontinuity in surface reactance does not seem to affect the excitation of radiation by the surface wave. The experiments have demonstrated that such a discontinuity has a very small power reflection coefficient. Further, the radiation excited by the surface wave at this type of discontinuity has been found to be more confined near the surface than that excited by the surface at the edge of a metallic strip above the surface.

The investigation indicates that powers radiated by the surface wave at a number of discontinuities can be controlled accurately. High-directive gain radiators can therefore be designed by having an array of discontinuities in the path of surface waves. From the point of view of minimizing the side lobes and powers carried by them, the use of an array of discontinuities in surface reactance seems to be preferable.

(6) ACKNOWLEDGMENTS

The work described was carried out at University College, London, as a part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research. The author wishes to express his gratitude to Professor H. E. M. Barlow for the research facilities provided and to Dr. J. Brown for the benefit of many stimulating discussions. He is also grateful to Mr. Effemey for helpful practical suggestions.

(7) REFERENCES

(1) Barlow, H. F. M., and Cullen, A. L.: 'Surface Waves', *Proceedings I.E.E.*, Paper No. 1482, April, 1953 (**100**, Part III, p. 329).

(2) Ehrlich, J., and Newkirk, L. L.: 'Corrugated Surface Antennas', *I.R.E. National Convention Record*, 1953, Part 2, p. 18.

(3) Jones, E. M. T.: 'An Annular Corrugated Surface Antenna', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 721.

(4) Brown, J., and Spector, J. O.: 'The Radiating Properties of End-Fire Aerials', *Proceedings I.E.E.*, Paper No. 2216 R, January, 1957 (**104** B, p. 27).

(5) Simon, J. C., and Weill, G.: 'A New Type of End-Fire Aerial', *Annales de Radioélectricité*, 1953, **8**, p. 183.

(6) Mueller, G. E.: 'A Broadside Dielectric Antenna', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 71.

(7) Deschamps, G. A.: 'Determination of Reflection Coefficient and Insertion Loss of a Wave-guide Junction', *Journal of Applied Physics*, 1953, **24**, p. 1046.

(8) Brown, J.: 'The Behaviour of Surface Waves at a Discontinuity in Surface Reactance' (To be published).

(9) Angulo, C. M.: 'Diffraction of Surface Waves by a Semi-Infinite Dielectric Slab', *Transactions of the Institute of Radio Engineers*, 1957, **AP5**, p. 100.

(10) Cullen, A. L.: 'The Excitation of Plane Surface Waves', *Proceedings I.E.E.*, Monograph No. 93, February, 1954 (**101**, Part IV, p. 225).

Cullen, A. L.: 'A Note on the Excitation of Plane Surface Waves', *ibid.*, Monograph No. 239 R, May, 1957 (**104** C, p. 472).

(11) Storer, J. E., Sheingold, L. S., and Stein, S.: 'A Simple Graphical Analysis of a Two Part Waveguide Junction', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 1004.

(12) Rich, G. J.: 'The Launching of a Plane Surface Wave', *Proceedings I.E.E.*, Paper No. 1783 R, March, 1955 (**102** B, p. 237).

(13) Discussion on Reference 12, *ibid.*, 1956, **103** B, p. 787.

(14) Fernando, W. M. G., and Barlow, H. E. M.: 'An Investigation of the Properties of Radial Cylindrical Surface Waves Launched over Flat Reactive Surfaces', *ibid.*, Paper No. 2009 R, May, 1956 (**103** B, p. 307).

621.372.826

The Institution of Electrical Engineers
Paper No. 2738 E
Mar. 1959

THE LAUNCHING OF RADIAL CYLINDRICAL SURFACE WAVES BY A CIRCUMFERENTIAL SLOT

By J. BROWN, M.A., Ph.D., Associate Member, and K. P. SHARMA, M.Sc., Ph.D., Graduate.

(*The paper was first received 26th April, and in revised form 23rd July*, 1958.)

SUMMARY

The efficiency with which radial cylindrical surface waves may be launched by a circumferential slot is investigated theoretically and experimentally. The radius of the slot has been found to affect the launching efficiency very little. An optimum efficiency of about 68% is obtainable with a slot of 2 cm radius above a 58-ohm reactive surface.

LIST OF PRINCIPAL SYMBOLS

E_y, E_r, E_ϕ = Components of electric field.
H_ϕ = Component of magnetic field.
k = Free-space phase-change coefficient.
X_s = Surface reactance of the guiding surface.
β_s = Surface-wave phase-change coefficient.
u = Surface-wave vertical decay coefficient.
μ_0 = Permeability of free space.
ϵ_0 = Permittivity of free space.
Z_0 = Free-space characteristic wave impedance.
H(. . .) = Hankel function.
r, ϕ, y = Cylindrical co-ordinates.
R, ϕ, θ = Spherical co-ordinates.
γ = Propagation coefficient in y-direction.
ω = Angular frequency.

(1) INTRODUCTION

Surface waves[1] are known to exist in a number of forms, and a considerable amount of work has been carried out on the efficiency with which such waves can be launched. Cullen[2] has shown that a suitably positioned slot can deliver about 95% of the available power to a plane surface wave supported by a purely reactive surface. His method of analysis has been extended by Fernando and Barlow[3] to give the efficiency with which a Hertzian doublet or a thin dipole can launch a radial cylindrical surface wave. Efficiencies in this case are rather lower, being around 80%. The practical launching system used by Fernando and Barlow[3] is a half-wave dipole fed by a coaxial line extended radially over the reactive surface which disturbs the circular symmetry of the surface wave. The present investigation was undertaken to excite a radial cylindrical surface wave of perfect circular symmetry, and also to determine whether a circumferential slot on a conducting cylinder can launch such a surface wave more efficiently than a dipole.

(2) FIELD EQUATIONS OF A RADIAL CYLINDRICAL SURFACE WAVE

The field components of a radial cylindrical surface wave are given by

$$E_y = A \exp(-uy)\mathrm{H}_0^{(2)}(\beta_s r) \quad . \quad . \quad . \quad . \quad (1)$$

$$E_r = \frac{Au}{\beta_s} \exp(-uy)\mathrm{H}_1^{(2)}(\beta_s r) \quad . \quad . \quad . \quad (2)$$

$$H_\phi = \frac{j\omega\epsilon_0 A}{\beta_s} \exp(-uy)\mathrm{H}_1^{(2)}(\beta_s r) \quad . \quad . \quad . \quad (3)$$

where

$$\beta_s^2 = k^2 + u^2 = \omega^2\mu_0\epsilon_0 + u^2 \quad . \quad . \quad . \quad (4)$$

If the surface is taken as the plane $y = 0$ (Fig. 1), then $-(E_r/H_\phi)_{y=0}$ is the surface reactance and has the value $(ju/\omega\epsilon_0) = jX_s$.

The time factor $\exp(j\omega t)$ has been suppressed in the field equations.

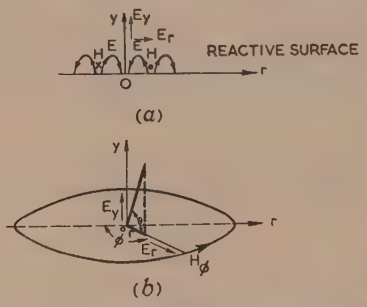


Fig. 1.—Radial cylindrical surface waves.
(*a*) Reactive surface supporting a radial cylindrical surface wave.
(*b*) Cylindrical co-ordinates and the field components.

(3) CALCULATION OF LAUNCHING EFFICIENCY

(3.1) General Considerations

The wave is to be launched by a circumferential slot in a long metallic cylinder over a reactive surface, and the problem with which we are concerned is the calculation of powers carried by the surface wave and by the radiation field when a voltage V is applied across the slot. This problem appears to be very similar to that discussed by Cullen,[2] the only difference being the change from plane to cylindrical geometry. A difficulty arises, however, in that the solution must be such that the electric-field component E_y vanishes over the surface of the conducting cylinder. In the case examined by Cullen,[2] the vanishing of E_y over the metallic sheet containing the slot follows automatically from the symmetry between the two sides of the slot. No similar result can be found in the present problem, and this leads to the need for an alternative method of analysis. A rigorous solution of the problem is given in Appendix 9. A physical argument which leads to the same answer is given in this Section.

It has been previously pointed out[4] that any form of surface-wave launcher will give rise not only to a surface wave but also to a continuous spectrum of plane waves which are partly of a propagating nature and partly evanescent. The fundamental properties of plane-wave spectra have been discussed in detail by Booker and Clemmow.[5] In the present problem the ampli-

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

tude of the surface wave can be readily determined by using a method suggested by Goubau,[6] and the function which determines the amplitudes of the plane waves in the spectrum can be obtained by the Booker–Clemmow[5] approach.

(3.2) Radiation Field

A rather more general case than the slot can be treated when it is supposed that the electric-field component E_y has a known value f(y) on the surface of the cylinder $r = a$ (Fig. 2), being

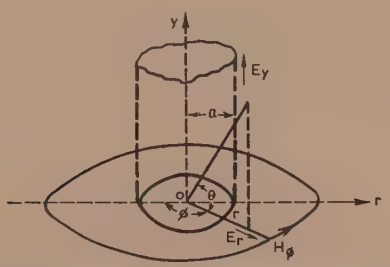


Fig. 2.—Cylindrical structure over a flat surface.
The surface of the cylinder has a specified distribution of E_y, and the flat surface is reactive.

defined only for y positive. Since E_y is independent of ϕ on the cylinder, the fields everywhere will be independent of ϕ; further, if E_ϕ is assumed identically zero on the cylinder $r = a$, it is easily shown that the only non-zero field components are E_y, E_r and H_ϕ.

The plane-wave spectrum must be selected to satisfy the condition that the plane $y = 0$ has the surface reactance given by

$$jX_s = -(E_r/H_\phi)_{y=0} = ju/\omega\epsilon_0 \quad . \quad . \quad . \quad (5)$$

This plane-wave spectrum arises from the presence of the specified electric field on the cylinder $r = a$, and must lead to a field which is symmetrical about the axis at $r = 0$. This enables the general spectrum, which involves integration over two variables, to be reduced to a spectrum of outward-travelling cylindrical waves, for which only one integral variable is needed. In particular, the component E_y can be written

$$E_y = \int_{-\infty}^{\infty} \mathrm{S}(p)\mathrm{H}_0^{(2)}[r\sqrt{(k^2 - p^2)}] \exp(-jpy)dp \quad . \quad (6)$$

where S(p) is the function which determines the amplitude of each wave in the spectrum. This function is determined to satisfy the prescribed form of E_y on the cylinder $r = a$. Suppose, for example, that the reactive surface is removed, that E_y has the value f(y) for positive values of y and is zero for negative values of y. Then, substitution of $r = a$ in eqn. (6) gives:

$$\int_{-\infty}^{\infty} \mathrm{S}(p)\mathrm{H}_0^{(2)}[a\sqrt{(k^2 - p^2)}] \exp(-jpy)dp = f(y) : y \geqslant 0$$
$$= 0 \quad : y \leqslant 0 \quad (7)$$

from which S(p) is given by taking Fourier transforms.

Let $$\mathrm{S}(p)\mathrm{H}_0^{(2)}[a\sqrt{(k^2 - p^2)}] = \mathrm{F}(p) \quad . \quad . \quad . \quad (8)$$

where $$\mathrm{F}(p) = \frac{1}{2\pi}\int_0^{\infty} \mathrm{f}(y) \exp(jpy)dy \quad . \quad . \quad . \quad . \quad (9)$$

S(p) is thus expressed in terms of known functions.

The next step is to determine the effect of the reactive surface in the plane $y = 0$. Each of the component waves in eqn. (6) will be incident upon the surface and generate a reflected wave. For example, the wave for which

$$E_y = \mathrm{S}(p)\mathrm{H}_0^{(2)}[r\sqrt{(k^2 - p^2)}] \exp(-jpy) \quad . \quad . \quad (10)$$

will set up a reflected wave for which

$$E_y = \mathrm{R}(p)\mathrm{S}(p)\mathrm{H}_0^{(2)}[r\sqrt{(k^2 - p^2)}] \exp(jpy) \quad . \quad (11)$$

R(p) having the significance of a reflection coefficient. This reflection coefficient can be calculated by imposing the condition that the sum of the incident and the reflected waves must satisfy eqn. (5). A straightforward calculation gives

$$\mathrm{R}(p) = (jp - u)/(jp + u) \quad . \quad . \quad . \quad . \quad . \quad (12)$$

The modulus of R(p) is unity, showing that no energy passes into the reactive surface.

Thus, when the reactive surface is introduced into the plane $y = 0$, the component E_y is given by

$$E_y = \int_{-\infty}^{\infty} \mathrm{S}(p)\mathrm{H}_0^{(2)}[r\sqrt{(k^2 - p^2)}][\exp(-jpy) + \mathrm{R}(p) \exp(jpy)]dp \quad . \quad . \quad . \quad . \quad (13)$$

where S(p) and R(p) are given by eqns. (8) and (12), respectively. Eqn. (13) gives the radiation field established by the specified field distribution on the cylinder $r = a$, and can be applied to the slot by substituting

$$\mathrm{f}(y) = V\delta(y - h) \quad . \quad . \quad . \quad . \quad . \quad (14)$$

which shows that f(y) is zero except at the slot position $y = h$, there being a voltage V across the slot. In this case

$$\mathrm{F}(p) = V \exp(jph)/2\pi \quad . \quad . \quad . \quad . \quad (15)$$

so that

$$E_y^R = \frac{V}{2\pi}\int_{-\infty}^{\infty} \frac{\mathrm{H}_0^{(2)}[r\sqrt{(k^2 - p^2)}]}{\mathrm{H}_0^{(2)}[a\sqrt{(k^2 - p^2)}]}\left\{\exp[-jp(y - h)] + \frac{jp - u}{jp + u}\exp(jp\overline{y + h})\right\}dp \quad (16)$$

(3.3) Surface-Wave Field

Goubau has shown[5] that the amplitudes of surface waves excited over purely reactive surfaces can be calculated from the field distribution which excites the waves. The result follows from the reciprocity theorem for electromagnetic fields. An alternative demonstration follows if it is assumed that the radiation and surface-wave fields are orthogonal in the sense that waveguide modes are orthogonal, i.e. the powers for the two can be calculated quite independently. This means that, if the known transverse electric-field distribution on the cylinder $r = a$ be multiplied by the magnetic-field component corresponding to the surface wave, and the result be integrated over the cylinder length, the result will not depend on the radiation field. Algebraically, this gives for the surface-wave amplitude, A,

$$\int_0^{\infty} \mathrm{f}(y)(j\omega\epsilon_0/\beta_s)A\mathrm{H}_1^{(2)}(\beta_s a) \exp(-uy)dy = \int_0^{\infty} (j\omega\epsilon_0/\beta_s)A^2\mathrm{H}_0^{(2)}(\beta_s a)\mathrm{H}_1^{(2)}(\beta_s a) \exp(-2uy)dy \quad (17)$$

since that part of f(y) corresponding to radiation field is orthogonal to exp ($-uy$). The evaluation of the integral on the right-hand side of eqn. (17) leads to

$$A = 2u\int_0^{\infty} \mathrm{f}(y) \exp(-uy)dy/\mathrm{H}_0^{(2)}(\beta_s a) \quad . \quad . \quad (18)$$

hich can be written as

$$A = 4\pi u \mathrm{F}(ju)/\mathrm{H}_0^{(2)}(\beta_s a) \quad . \quad . \quad . \quad . \quad (19)$$

y using eqn. (9).
or the slot,

$$\mathrm{F}(ju) = V \exp(-uh)/2\pi \quad . \quad . \quad . \quad . \quad (20)$$

o that the surface-wave field established by the slot is given by

$$E_y^S = 2uV \exp[-u(y+h)]\mathrm{H}_0^{(2)}(\beta_s r)/\mathrm{H}_0^{(2)}(\beta_s a) \quad (21)$$

(3.4) Power in the Radiation Field

The power in the radiation field is calculated from the electric-eld component E_θ^R and need only be evaluated at a large istance from the launcher. The Hankel function in the integral quation giving the field can then be replaced by the first term n its asymptotic expansion. The integral has been evaluated y the method of steepest descent to give

$$E_R = -\frac{jV \sec\theta \mathrm{P}(\theta) \exp(-jkR)}{\pi R \mathrm{H}_0^{(2)}(ka\cos\theta)} \quad . \quad . \quad (22)$$

here

$$\mathrm{P}(\theta) = \left[\exp(jkh\sin\theta) + \frac{jk\sin\theta - u}{jk\sin\theta + u}\exp(-jkh\sin\theta)\right] \quad (23)$$

nd R, θ are spherical polar co-ordinates, so that the radiated ower is

$$P_R = \frac{V^2}{Z_0\pi^2}\int_0^{\pi/2} \frac{|P(\theta)|^2 \sec\theta d\theta}{|\mathrm{H}_0^{(2)}(ka\cos\theta)|^2} \quad . \quad . \quad . \quad (24)$$

(3.5) Power in the Surface-Wave Field

The power in the surface wave has been calculated by inte-rating the Poynting vector over a cylindrical surface coaxial ith the axis at $r = 0$. The final expression is

$$P_S = \frac{4V^2 u\omega\epsilon_0 \exp(-2uh)}{\beta_s^2 |\mathrm{H}_0^{(2)}(\beta_s a)|^2} \quad . \quad . \quad . \quad . \quad (25)$$

(3.6) Launching Efficiency

By definition, the launching efficiency is

$$\eta = \frac{P_S}{P_S + P_R} = \frac{4uk\exp(-2uh)}{4uk\exp(-2uh) + \beta_s^2|\mathrm{H}_0^{(2)}(\beta_s a)|^2 I} \quad (26)$$

here

$$I = \int_0^{\pi/2} \frac{|\mathrm{P}(\theta)|^2 \sec\theta d\theta}{|\mathrm{H}_0^{(2)}(ka\cos\theta)|^2} \quad . \quad . \quad . \quad . \quad (27)$$

he integral in eqn. (27) can be evaluated numerically.

The limiting case of a tending to infinity corresponds to an nfinitely long horizontal-slot source over a reactive surface, lready examined by Cullen.[2] It is readily verified that his result grees with the limit of eqn. (26) as a tends to infinity.

Eqn. (26) has been used to calculate the launching efficiency f the surface wave supported by a 58-ohm reactive surface at frequency of 9·375 Gc/s. The results for $a = 1$ cm, 2 cm nd infinity are shown in Fig. 3. The launching efficiency seems o vary very little with the slot radius. A slot of 2 cm radius t a height of 0·8 cm above the 58-ohm reactive surface provides maximum launching efficiency of 68%.

Cullen[2] has shown that an infinitely long horizontal slot at a eight of about $0{\cdot}63/u$ cm over a corrugated surface of $u/k = 0{\cdot}5$, an provide a launching efficiency of 95·2%. Calculation shows hat, under similar conditions, a circumferential slot of 2 cm adius can provide a launching efficiency of 94·3%.

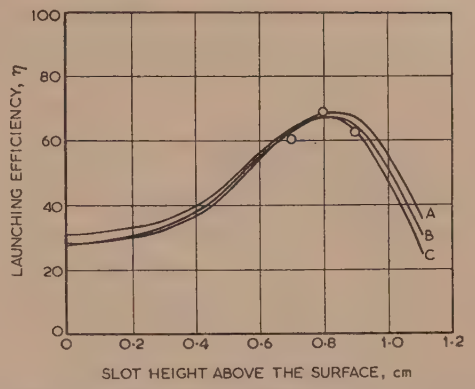


Fig. 3.—Variation of launching efficiency of a circumferential slot with the radius of the slot, and with its height above the corrugated surface.

Frequency = 9·375 Gc/s.
A: $a \to \infty$.
B: $a = 2$ cm.
C: $a = 1$ cm.
⊙ Experimental points for $a = 2$ cm.

(4) EXPERIMENTAL WORK

The launcher used is shown in Fig. 4 and consists of a circumferential slot in a conducting cylinder. The slot is at the circumference of a radial line fed by a coaxial line within the cylinder. Ideally the height of the cylinder should be infinite, but, in practice, the surface-wave decay coefficient is so rapid that a height of about 5 in is adequate.

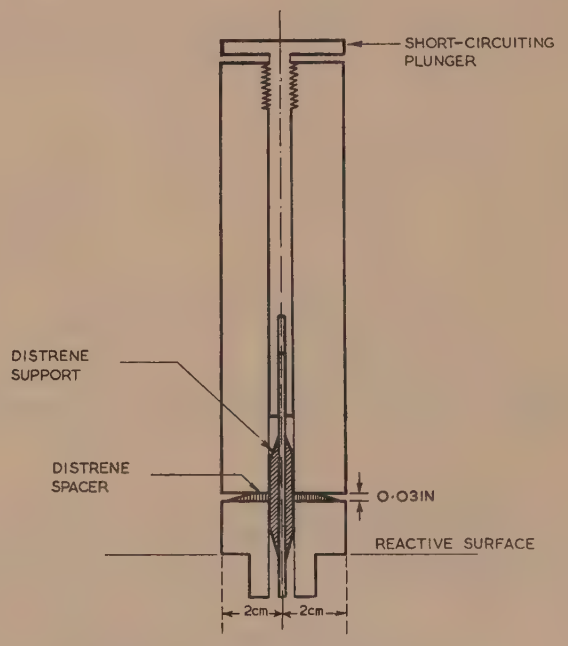


Fig. 4.—Circumferential slot launcher.

The reactive surface which was used in the experiments is the corrugated aluminium plate described by Fernando and Barlow,[3] the surface reactance being 58 ohms at 9·375 Gc/s. The plate is surrounded by a resistive load, formed from carbon-coated strawboard, which provides a satisfactory match for the surface wave. The vertical field distribution, from which the power in the surface wave is calculated, is measured as described by Fernando and Barlow.[3] The radiated power is obtained by integration of the radiation pattern, which is measured by equipment similar to that described by Brick.[7]

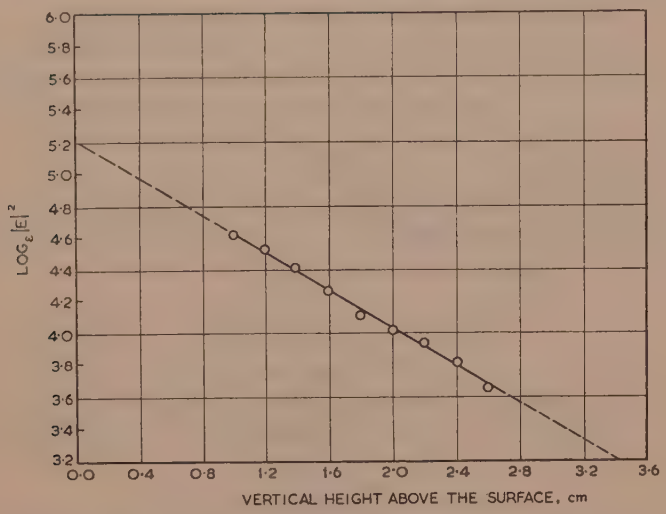


Fig. 5.—Logarithm of the power-flow density plotted against the height above the surface.
Frequency = 9·375 Gc/s. $k_1 = 0·293$

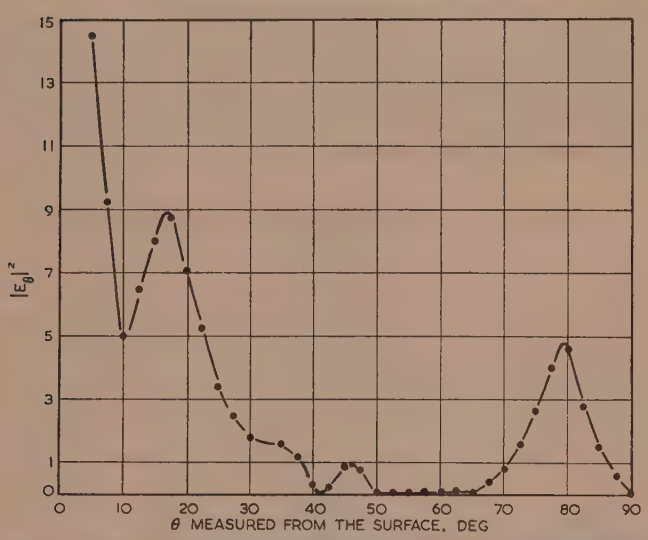


Fig. 6.—Polar power-flow-density distribution for a circumferential slot over a corrugated surface, measured at a radial distance of 53 cm.
Height of slot = 0·8 cm. Frequency = 9375 Mc/s.

Experimental results are shown in Figs. 5 and 6. The former shows the logarithm of the power-flow density plotted against the height above the surface; the linear relationship confirms that a surface wave is excited, and the slope of the line agrees with the theoretical value corresponding to the 58-ohm surface reactance. Fig. 6 gives the measured power radiation pattern when the height of the launching slot is 0·8 cm, the corresponding calculated pattern being shown in Fig. 7. The peak at the left of Fig. 6, i.e. near zero, arises from the surface wave and is therefore not part of the radiation pattern. The rapid decay of the radiation field for $\theta > 80°$ is probably dependent on the height of the conducting cylinder. In any case, the proportion of the total radiated power which emerges in the range 80–90° is small because of the small solid angle subtended at the launcher. Figs. 6 and 7 agree in the positions of the peak (at about 17°) and the null (at 50°). There is therefore reasonable agreement between the experimental and theoretical values.

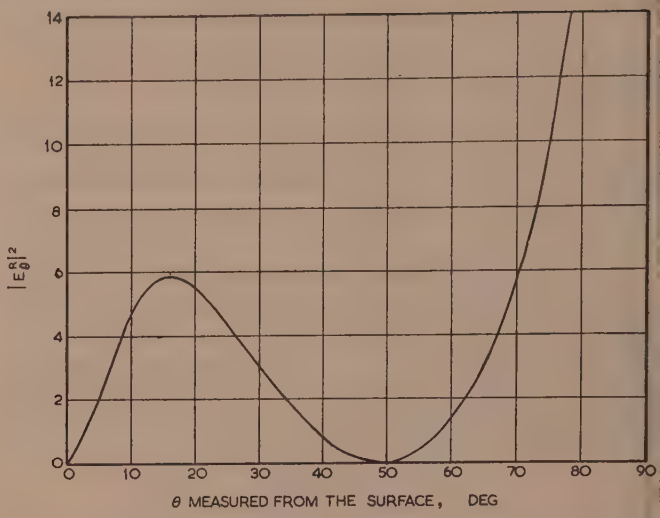


Fig. 7.—Theoretical radiation-field power-flow-density pattern of a circumferential-slot launcher.
Data for launcher as in Fig. 6.

The launching efficiencies deduced from the above experimental results are shown in Fig. 3, and agree with the theoretical figures, obtained in previous Sections, within the limits of experimental error.

(5) THE HALF-WAVE DIPOLE LAUNCHER

Fernando and Barlow[3] described one form of half-wave dipole launcher, with top and bottom halves of equal thickness, this thickness being appreciable compared with the dipole length. The launching efficiency of this arrangement was found to be poor, and one reason for the investigation described in the paper was the desire to improve the efficiency. The Fernando and Barlow launcher approximates to a circumferential slot in a short metallic cylinder, and it would be expected to have an efficiency of about 60%, corresponding to the slot $\lambda_0/4$ above the surface (Fig. 3). In practice, it differs from the arrangement discussed in the present paper in two respects:

(*a*) The lower limb of the dipole does not make contact with the reactive surface.
(*b*) The upper limb is only $\lambda_0/4$ long, whereas the cylinder is assumed to be infinite in the theoretical analysis.

As a result of these two differences it is probable that the current distribution on the walls of the Fernando–Barlow dipole differs appreciably from that on the ideal cylinder, and this could explain the observed poor launching efficiency.

The dipole has the advantage of small size compared with the slot launcher of Fig. 4, and an attempt was made to achieve a design which would be efficient. Point (*a*) is easily met, and the effects of the current distribution on the upper half can be altered by changing the radius. The simplest possibility is to make the upper half of the same radius as the inner of the coaxial feed. The launching efficiency of this dipole was found by the method of the previous Section to be 73%, and by an alternative method[8] to be 71%. This is rather higher than the maximum value in Fig. 3. The explanation is that the contributions to the surface wave from the currents in the different parts of the dipole reinforce, the dipole therefore being equivalent to a distributed launcher. A corresponding gain in efficiency could be achieved by using a wide slot or a series of narrow slots, but in view of the complexity, this has not been considered worth pursuing.

(6) CONCLUSIONS

Calculations and measurements show that a circumferential lot in a conducting cylinder can act as an efficient launcher of radial surface wave having perfect cylindrical symmetry, and hat the experimental results are in agreement with theory. The investigation of this arrangement has led to an understanding of the half-wave dipole launcher and to a modified arrangement of higher launching efficiency.

(7) ACKNOWLEDGMENTS

The work described in the paper was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research. The authors are indebted to Prof. H. E. M. Barlow and Mr. H. G. Effemey for many helpful discussions.

(8) REFERENCES

1) BARLOW, H. E. M., and CULLEN, A. L.: 'Surface Waves', *Proceedings I.E.E.*, Paper No. 1482, April, 1953 (**100**, Part III, p. 329).
2) CULLEN, A. L.: 'The Excitation of Plane Surface Waves', *ibid.*, Monograph No. 93, February, 1954 (**101**, Part IV, p. 225).
3) FERNANDO, W. M. G., and BARLOW, H. E. M.: 'An Investigation on the Properties of Radial Cylindrical Surface Waves launched over Flat Reactive Surfaces,' *ibid.*, Paper No. 2009 R, May, 1956 (**103** B, p. 307).
4) BROWN, J.: 'The Types of Wave which may exist near a Guiding Surface', *ibid.*, Paper No. 1567, November, 1953 (**100**, Part III, p. 363).
5) BOOKER, H. G., and CLEMMOW, P. C.: 'The Concept of an Angular Spectrum of Plane Waves, and its Relation to that of Polar Diagram and Aperture Distribution', *ibid.*, Paper No. 922, January, 1950 (**97**, Part III, p. 11).
6) GOUBAU, G.: 'On the Excitation of Surface Waves', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 865.
7) BRICK, D. B.: 'The Excitation of Surface Waves by a Vertical Antenna', *ibid.*, 1955, **43**, p. 721.
8) SHARMA, K. P.: 'An Investigation of the Excitation of Radiation by Surface Waves' (see page 116).

(9) APPENDIX

(9.1) The Field Excited by a Specified Field Distribution over a Semi-Infinite Cylinder above a Reactive Surface

The problem is to determine a solution of Maxwell's equations in the region $y \geqslant 0$, $r \geqslant a$, given that the tangential electric field on the cylinder $r = a$ is

$$E_y = \mathrm{f}(y) : y \geqslant 0 \quad . \quad . \quad . \quad . \quad . \quad (28)$$

and that the reactance of the plane $y = 0$ is $u/\omega\epsilon_0$. Since E_y is independent of ϕ and since E_ϕ is identically zero on the cylinder $r = a$, the only non-zero field components are E_y, E_r and H_ϕ. These are related by

$$j\omega\epsilon_0 E_r = -\frac{\partial H_\phi}{\partial y} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

$$j\omega\epsilon_0 E_y = \frac{1}{r}\frac{\partial}{\partial r}(rH_\phi) \quad . \quad . \quad . \quad . \quad (30)$$

The field must correspond to a radiation field for large values of r. Consider, therefore, the field given by

$$H_\phi = \sin(py + \psi)\mathrm{H}_1^{(2)}[r\sqrt{(k^2 - p^2)}] \quad . \quad . \quad (31)$$

which satisfies the wave equation and the radiation condition for large values of r. It will also satisfy the condition that the reactance of the plane $y = 0$ is $u/\omega\epsilon_0$, provided that

$$\left[\frac{E_r}{H_\phi}\right]_{y=0} = -\frac{p\cos\psi}{j\omega\epsilon_0 \sin\psi} = -\frac{ju}{\omega\epsilon_0}$$

i.e.
$$\tan\psi = -p/u \quad . \quad . \quad . \quad . \quad . \quad (32)$$

Any solution in the required region can therefore be represented by

$$H_\phi = \int_0^\infty \mathrm{A}(p)\sin(py + \psi)\mathrm{H}_1^{(2)}[r\sqrt{(k^2 - p^2)}]dp \quad . \quad (33)$$

where $\mathrm{A}(p)$ is a suitably chosen function. The corresponding expression for the y-component of electric field is

$$E_y = \frac{1}{j\omega\epsilon_0}\int_0^\infty \mathrm{A}(p)\sin(py + \psi)\sqrt{(k^2 - p^2)}\mathrm{H}_0^{(2)}[\sqrt{(k^2 - p^2)}r]dp \quad . \quad . \quad . \quad . \quad (34)$$

since

$$\frac{1}{r}\frac{d}{dr}\{r\mathrm{H}_1^{(2)}[\sqrt{(k^2 - p^2)}r]\} = \sqrt{(k^2 - p^2)}\mathrm{H}_0^{(2)}[\sqrt{(k^2 - p^2)}r] \quad . \quad . \quad . \quad . \quad (35)$$

Eqn. (34) can be rewritten

$$E_y = -\frac{1}{2\omega\epsilon_0}\int_0^\infty \mathrm{A}(p)\sqrt{(k^2 - p^2)}[\exp(jpy + j\psi) - \exp(-jpy - j\psi)]\mathrm{H}_0^{(2)}[(\sqrt{k^2 - p^2})r]dp \quad (36)$$

Now $\sqrt{(k^2 - p^2)}$ is an even function of p and ψ is an odd function of p [see eqn. (32)], so that

$$E_y = -\frac{1}{2\omega\epsilon_0}\int_{-\infty}^\infty \mathrm{A}(p)\sqrt{(k^2 - p^2)}\exp(jpy + j\psi)\mathrm{H}_0^{(2)}[\sqrt{(k^2 - p^2)}r]dp \quad (37)$$

provided that
$$\mathrm{A}(-p) = -\mathrm{A}(p) \quad . \quad . \quad . \quad . \quad (38)$$

When $r = a$,

$$E_y = -\frac{1}{2\omega\epsilon_0}\int_{-\infty}^\infty \mathrm{A}(p)\sqrt{(k^2 - p^2)}\exp(jpy + j\psi)\mathrm{H}_0^{(2)}[\sqrt{(k^2 - p^2)}a]dp \quad (39)$$

Now, E_y has the known value $\mathrm{f}(y)$ for $y \geqslant 0$ and $r = a$: suppose that E_y as defined by eqn. (39) equals $\mathrm{h}(y)$ for $y \leqslant 0$, Then

$$\left.\begin{aligned}\int_{-\infty}^\infty \mathrm{C}(p)\exp(jpy + j\psi)dp &= \mathrm{f}(y) : y \geqslant 0\\ &= \mathrm{h}(y) \quad y \leqslant 0\end{aligned}\right\} \quad . \quad (40)$$

where
$$\mathrm{C}(p) = -\frac{1}{2\omega\epsilon_0}\sqrt{(k^2 - p^2)}\mathrm{H}_0^{(2)}[\sqrt{(k^2 - p^2)}a]\mathrm{A}(p) \quad (41)$$

and
$$\mathrm{C}(p) = -\mathrm{C}(-p) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

as a result of eqns. (38) and (41).

Take the Fourier transforms of both sides of eqn. (40).

Then
$$\mathrm{C}(p)\exp(j\psi) = \mathrm{F}(-p) + \mathrm{H}(p) \quad . \quad . \quad . \quad (43)$$

where $\mathrm{F}(p)$ is defined by eqn. (9)

and
$$\mathrm{H}(p) = \frac{1}{2\pi}\int_{-\infty}^0 \mathrm{h}(y)\exp(-jpy)dy \quad . \quad . \quad . \quad (44)$$

From the possible forms of the functions $\mathrm{f}(y)$ and $\mathrm{h}(y)$ it can be shown that $\mathrm{F}(-p)$ will be regular for $\mathscr{I}p < \delta$ and $\mathrm{H}(p)$ will be regular for $\mathscr{I}p > -\delta$, where δ is some positive quantity.

From eqns. (38) and (43),

$$C(p)\exp(-j\psi) = -F(+p) - H(-p) \quad . \quad (45)$$

and from eqn. (32),

$$\exp(j\psi) = \cos\psi + j\sin\psi = (u - jp/u + jp)^{1/2} \quad (46)$$

Elimination of $C(p)$ from eqns. (43) and (45) therefore leads to

$$-F(+p) - H(-p) = \frac{u + jp}{u - jp}[F(-p) + H(p)] \quad . \quad (47)$$

This equation is now rewritten in a form such that the left-hand side involves only functions which are regular for $\mathscr{I}p < \delta$ and the right-side only functions which are regular for $\mathscr{I}p > -\delta$. This is a step which occurs in the Wiener–Hopf method for solving integral equations and can be easily effected, with the following result:

$$\frac{F(-p) - F(+ju)}{u - jp} + \frac{H(-p)}{u + jp} + \frac{F(+ju)}{u + jp} = -\frac{F(+p) - F(+ju)}{u + jp} - \frac{H(p)}{u - jp} - \frac{F(+ju)}{u - jp} \quad (48)$$

The two sides of this equation have the common strip of regularity $-\delta < \mathscr{I}p < \delta$, and this condition, together with the behaviour of the functions as p tends to infinity, can be used to show that both sides must be zero. Hence

$$H(p) = \frac{jp - u}{jp + u}[F(+p) - F(+ju)] - F(+ju) = \frac{jp - u}{jp + u}F(+p) - \frac{2jp}{jp + u}F(+ju) \quad . \quad . \quad (49)$$

The function $C(p)$ can now be determined from eqn. (43), whence $A(p)$ can be found and the result substituted in eqn. (37) to give:

$$E_y = \int_{-\infty}^{\infty}\left\{F(-p) + \frac{jp - u}{jp + u}F(p) - \frac{2jp}{jp + u}F(ju)\right\}\exp(jpy)\frac{H_0^{(2)}[\sqrt{(k^2 - p^2)}r]}{H_0^{(2)}[\sqrt{(k^2 - p^2)}a]}dp \quad (50)$$

Consider this integral in three parts, corresponding to the three terms within the curly brackets: the second is left unaltered, in the first the integration variable is changed from p to $-p$ and the third is evaluated by residues, giving

$$E_y = \int_{-\infty}^{\infty}F(p)\left\{\exp(-jpy) + \frac{jp - u}{jp + u}\exp(jpy)\right\}\frac{H_0^{(2)}[\sqrt{(k^2 - p^2)}r]}{H_0^{(2)}[\sqrt{(k^2 - p^2)}a]}dp + 4\pi u F(ju)\exp(-uy)\frac{H_0^{(2)}(\beta_s r)}{H_0^{(2)}(\beta_s a)} \quad (51)$$

which proves the results deduced from physical arguments in Sections 3.2 and 3.3.

621.372.821

The Institution of Electrical Engineers
Paper No. 2770 E
Mar. 1959
©

PARALLEL-PLATE TRANSMISSION SYSTEMS FOR MICROWAVE FREQUENCIES

By A. F. HARVEY, D.Phil., B.Sc.(Eng.), Member.

(*The paper was first received* 23*rd April, in revised form* 24*th July, and in final form* 21*st August*, 1958.)

SUMMARY

The paper gives a practical account of parallel-plate, or strip, transmission lines, which are becoming of increasing importance in microwave applications. The different types of strip line are considered and their basic characteristics summarized. Transducers to conventional systems and various circuit-elements are described, while the application of strip-line techniques to the construction of microwave components and filters is examined. Manufacture by photo-etching is outlined.

(1) INTRODUCTION

The planar, or parallel-plate line, may be considered as an evolution[6] of the coaxial line. Thus a circular-section coaxial line may be deformed in such a manner that both the centre and outer conductors are first square and then rectangular. The narrow side walls are then taken to infinity to give a flat-strip transmission system. This is a symmetrical line, but by removing one of the outer plates an unsymmetrical system is obtained. Although Park[47] has described a balanced line consisting of two parallel strips placed between, and perpendicular to, two ground planes, most practical strip lines are unbalanced electrically.

Parallel-plate lines may be supported by a solid dielectric or air-spaced. An unsymmetrical or single-ground-plane dielectric-supported line is shown in Fig. 1(*a*), while a symmetrical, triplate or double-ground-plane line is shown in Fig. 1(*b*). The conductors are usually of copper and the dielectric is a low-loss material such as p.t.f.e. or silicone-impregnated Fibreglass. Lower attenuation is achieved with air-spaced lines, a simple strip above ground giving an unsymmetrical line. Three types of air-spaced triplate line are shown in Fig. 2, from which it will be seen that the amount of solid dielectric is reduced to the minimum required to give mechanical support to the inner conductor.

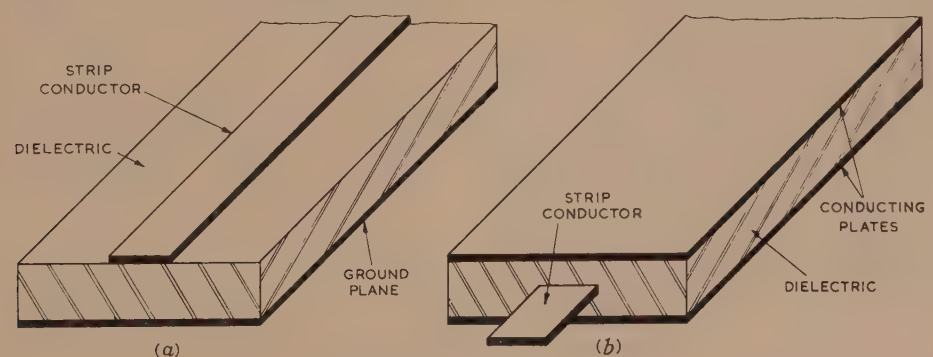


Fig. 1.—Types of dielectric-supported strip-line.
(*a*) The unsymmetrical line or microstrip.
(*b*) The symmetrical or sandwich line.

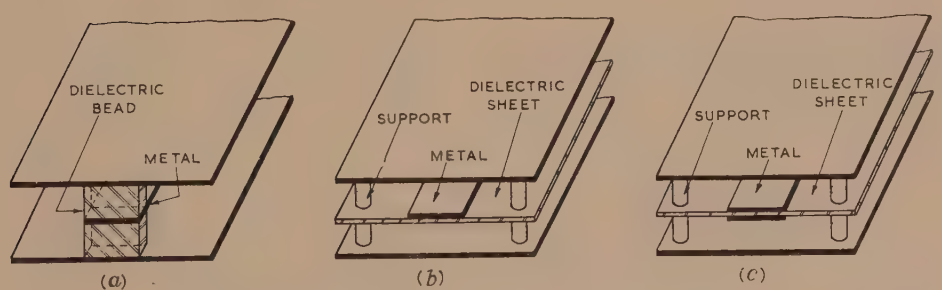


Fig. 2.—Types of high-Q-factor strip transmission line.
(*a*) With thin dielectric beads.
(*b*) With a thin dielectric sheet, single metalclad.
(*c*) With a thin dielectric sheet, double metalclad.

Despite the physical simplicity of these transmission systems, the rigorous analysis of their properties presents difficulty. The presence, in some instances, of a dielectric with a longitudinal interface necessitates the existence of longitudinal electric and magnetic fields. Thus a TEM-mode solution is inapplicable except as an approximation, and the nature of the dominant

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
This in an 'integrating' paper. Members are invited to submit papers in this category, giving the full perspective of the developments leading to the present practice in a particular part of one of the branches of electrical science.
Dr. Harvey is at the Royal Radar Establishment.

mode must be determined by direct application of Maxwell's equations for the particular boundary conditions chosen.

In such systems the usual circuit-elements can be constructed and counterparts of most coaxial-line and waveguide components may be realized. They are capable of broad-band performance, and complicated circuits can be readily constructed. Parallel-plate lines have additional advantages of relatively small bulk and weight, while they are capable of economic manufacture in quantity by printing techniques. The present state of development is, however, such that adaptors are necessary to accommodate associated items such as semiconductor diodes and electron tubes, while some components tend to be expensive and cumbersome. Parallel-plate lines also have to compete, for broad-band performance, with ridged waveguides, but nevertheless, they are finding increasing application in the frequency range 1–12 Gc/s.

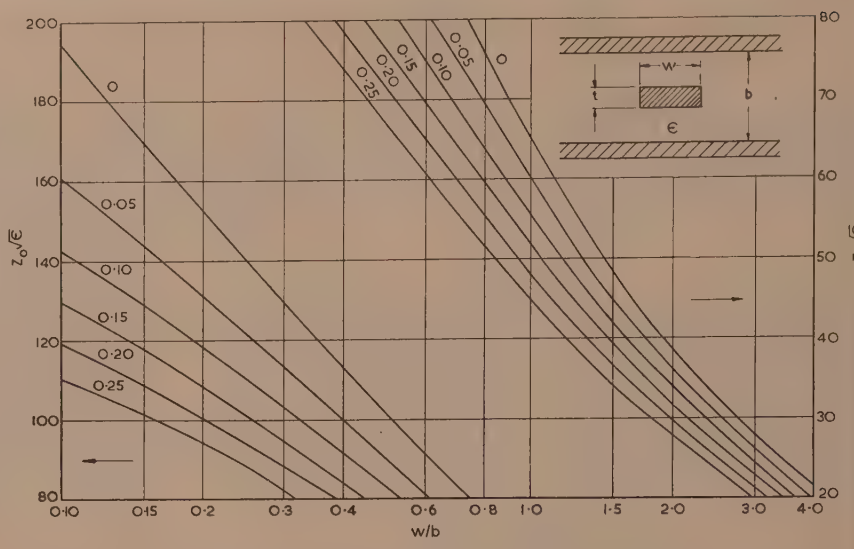


Fig. 3.—Characteristic impedance of a shielded strip-line.

Z_0 as a function of w/b: the numbers are values of t/b.

(2) DOUBLE-GROUND-PLANE LINES

(2.1) Solid-Dielectric Sandwich

The transmission system consisting of a thin metallic strip placed between two sheets of low-loss solid dielectric, of which the outer sides are covered with conducting planes, was first described by Barrett and Barnes.[7] In the theoretical analysis of Hayt,[37] Begovitch[8] and Dahlman[18] it was shown that, if the width of the outer strips is large compared with that of the inner, the electric field does not extend outside the dielectric and propagation is in the TEM-mode. The phase velocity in the strip line is then $1/\sqrt{\epsilon}$ times the free-space value, where ϵ is the dielectric constant of the line material, giving a guide wavelength of

$$\lambda_g = \lambda/\sqrt{\epsilon} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where λ is the free-space wavelength.

The characteristic impedance, Z_0, is given in ohms by

$$Z_0 = 10^8(\mu\epsilon)^{1/2}/3C \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where μ, the relative permeability, is unity for most dielectrics, and C is the capacitance per unit length in picofarads per metre.

This capacitance may be calculated by a conformal mapping method to give a characteristic impedance, assuming a central strip of zero thickness, of

$$Z_0 = 30\pi \mathrm{K}(k)/\mathrm{K}(1-k^2)^{1/2} \quad . \quad . \quad . \quad . \quad (3)$$

where $k = 1/\cosh(\pi w/2b)$, w is the strip width, b is the ground-plane spacing and $\mathrm{K}(k)$ is the complete elliptic integral of the first kind of modulus k. When $(w/b) \gg 1$,

$$Z_0 = 15\pi^2/[(\pi w/2b) + \log_\epsilon 2] \quad . \quad . \quad . \quad . \quad (4)$$

and when $(w/b) \ll 1$,

$$Z_0 = 60 \log_\epsilon (8b/\pi w) \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Barrett[6] found good agreement between these relations and experimental results at low frequencies. This analysis was extended by Pease and Mingins[49] and Cohn[14, 15] to take into account the finite thickness, t, of the inner conductor. Graphs of the characteristic impedance versus w/b for various values of t/b are shown in Fig. 3.

The attenuation due to dielectric loss is given in decibels per unit length, by

$$\alpha_d = (27{\cdot}3\sqrt{\epsilon}\tan\delta)/\lambda \quad . \quad . \quad . \quad . \quad . \quad (6)$$

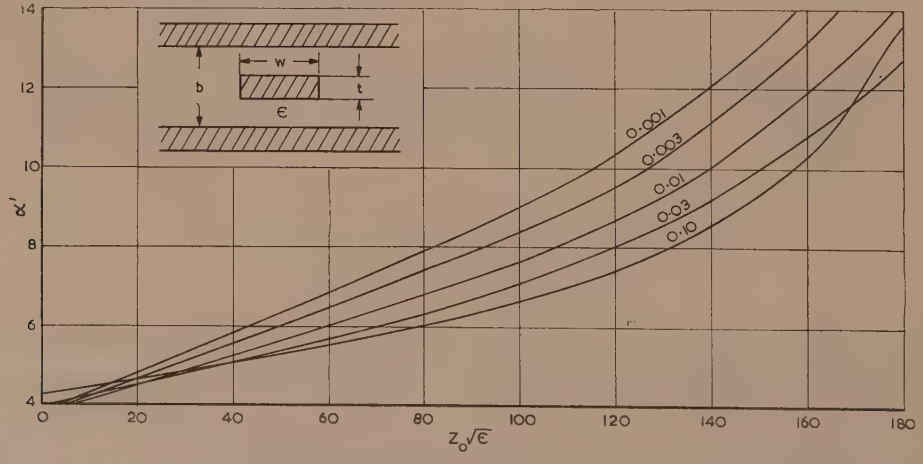


Fig. 4.—Attenuation of a shielded strip-line.

The curves are for copper and different values of t/b. $\alpha' = \dfrac{\alpha_c b \times 10^4}{\sqrt{(f\sigma_{c/s}\epsilon)}}$ dB/unit length.

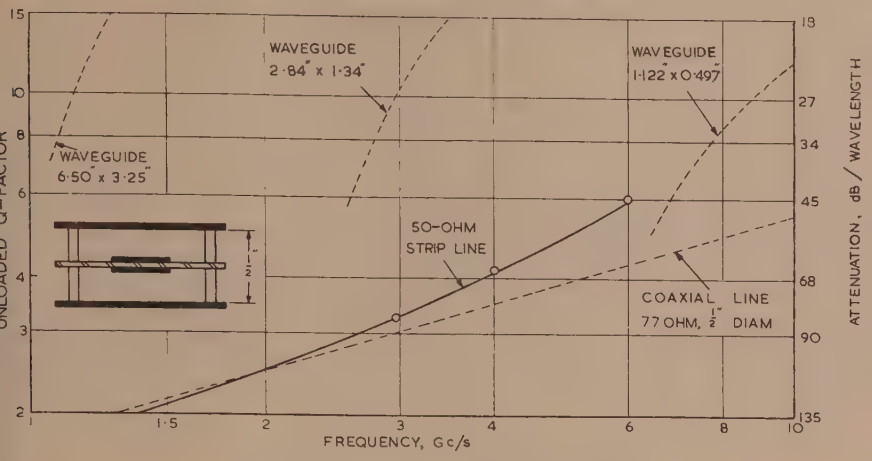


Fig. 5.—Attenuation of triplate strip line.

The unloaded Q-factor, Q_u, and attenuation per wavelength, α, are related by $Q_u\alpha = 27$. Some figures for coaxial line and waveguides are shown for comparison. All conductors are silver.

where tan δ is the loss tangent of the medium. The calculation of conductor-loss attenuation, α_c, is more difficult, but Cohn[15] derived exact formulae for both wide and narrow strips and fitted a reasonable transition so that the composite curve is correct to within a few per cent. Resultant graphs, for t/b ratios from 0·001 to 0·1, are shown in Fig. 4, thus enabling the attenuation to be obtained by simple calculation. For example, with polystyrene ($\epsilon = 2{\cdot}5$, $\tan\delta = 0{\cdot}001$) as dielectric and $w = 0{\cdot}15$ in, $b = 0{\cdot}125$ in, $t = 0{\cdot}001\,35$ in, then $Z_0 = 50$ ohms, $\alpha_d = 0{\cdot}43$ dB/m and $\alpha_c = 0{\cdot}83$ dB/m.

Fig. 4 shows that the conductor-loss attenuation decreases as Z_0 approaches zero, a condition which Fig. 3 associates with a strip of infinite width. If field fringing, with consequent non-uniformity of current distribution, did not occur, the attenuation would be independent of strip width. The effect of current non-uniformity on attenuation is therefore quite large in the useful range of characteristic impedance. If the dielectric material fills the line, the total attenuation is given, in decibels per unit length, by

$$\alpha = \alpha_c + (27{\cdot}3\sqrt{\epsilon}^{1/2}\tan\delta)/\lambda \quad . \quad . \quad . \quad . \quad (7)$$

To ensure that only the dominant TEM-mode can exist, the spacing between the strip conductor and the adjacent ground plane must be less than a quarter wavelength in the dielectric. Where small radiation loss is desirable the strip conductor must be midway between the ground planes and parallel to them. Likewise the equivalent electrical width of the central conductor must be rather less than half a wavelength; a more precise condition has, however, been determined by Oliner.[45]

(2.2) Air-Spaced Triplate Line

An air-spaced construction practically eliminates dielectric losses and gives what is generally known as the high-Q-factor triplate line. Fromm[29] has measured the unloaded Q-factor, Q_u, of such a line, the dimensions and results being given in Fig. 5. The attenuation, in decibels per wavelength, calculated from the expression

$$\alpha = 27/Q_u \quad . \quad . \quad . \quad . \quad (8)$$

is also shown. The Q-factor is, in general, quite high and rises as either the ground plane spacing or the width of the strip is increased. Q-factors some 50% higher may be obtained by increasing the strip width beyond an electrical half-wavelength, the exact value being chosen so that transverse resonances are avoided. For comparison, Fig. 5 also shows the computed attenuation for an air-spaced coaxial line ½ in in diameter, using the characteristic impedance of 77 ohms, which has the highest Q-factor, together with the computed attenuations for rectangular TE_{01}-mode waveguides covering the same frequency range. The attenuation of the strip line is lower than that of the coaxial line but higher than that of the waveguides, which are, however, inferior in bandwidth.

(3) SINGLE-GROUND-PLANE LINES

(3.1) Air-Spaced Strip-Above-Ground Lines

A simple transmission line consists of an asymmetrical air-spaced conductor above ground, the conductor being a round rod[35] or rectangular strip.[62] For a strip of width w, negligible thickness t, and spaced distance b from the ground plane, Assadourian and Rimai[4] gave an approximate expression, with w/b large, for the characteristic impedance as

$$Z_0 = \frac{Z}{\left\{\left(\frac{w}{b}\right) + \left(\frac{2}{\pi}\right)\left[1 + \log_\epsilon\left(1 + \frac{\pi w}{2b}\right)\right]\right\}} \quad . \quad (9)$$

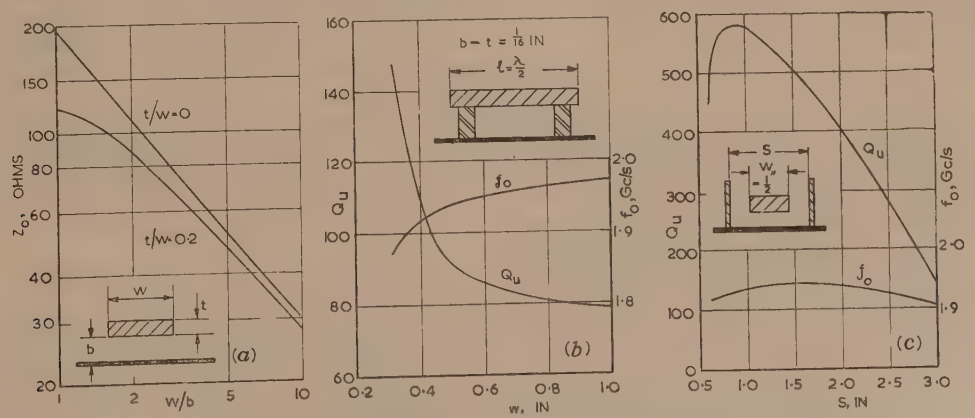


Fig. 6.—Properties of asymmetrical strip-above-ground.

(*a*) Characteristic impedance with t/w as parameter.
(*b*) Unloaded Q-factor and resonant frequency.
(*c*) Improvement due to shielding.

where $Z = \sqrt{(\mu_0/\epsilon_0)}$ is the characteristic impedance of free space. This relation is shown by the straight line in Fig. 6(*a*).

For t finite but small, the characteristic impedance becomes

$$Z_0 = \frac{Z}{\left\{2 + \left(\frac{w}{b}\right) + \left(\frac{t}{\pi b}\right)\left[1 + \log_e\left(1 + \frac{2b}{t}\right)\right]\right\}} \quad . \quad (10)$$

which is shown, for the case of $t/w = 0{\cdot}2$, by the curved line of Fig. 6(*a*). The characteristic impedance is seen to be controlled by the ratio w/b and, to a smaller extent, by the parameter ratio t/w. An impedance of, for example, 50 ohms is achieved with $w/b = 5$ and $t/w = 0{\cdot}2$: practical values for a line operating at a frequency of 2 Gc/s are $w = \frac{5}{16}$ in and $b = t = \frac{1}{16}$ in.

Zublin[62] examined the properties of such a strip line, one half-wavelength long and short-circuited at both ends, as a resonator. The half-power bandwidth was measured for various strip widths and spacings from the ground plane. For a constant spacing of $\frac{1}{16}$ in, Fig. 6(*b*) represents the unloaded Q-factor as a function of strip width: the resonant frequency of the structure is also shown. With smaller spacing, the Q-factor was found to increase and the resonant frequency to decrease. The comparatively low Q-factors are due to strong radiation from the ends of the resonator. This radiation was reduced by means of two shields, 2 in high by 5 in long, parallel to the resonant strip and symmetrically placed. The improvement in Q-factor so obtained is shown by the results in Fig. 6(*c*): for the optimum shield spacing the Q-factor increased from 96 to 580.

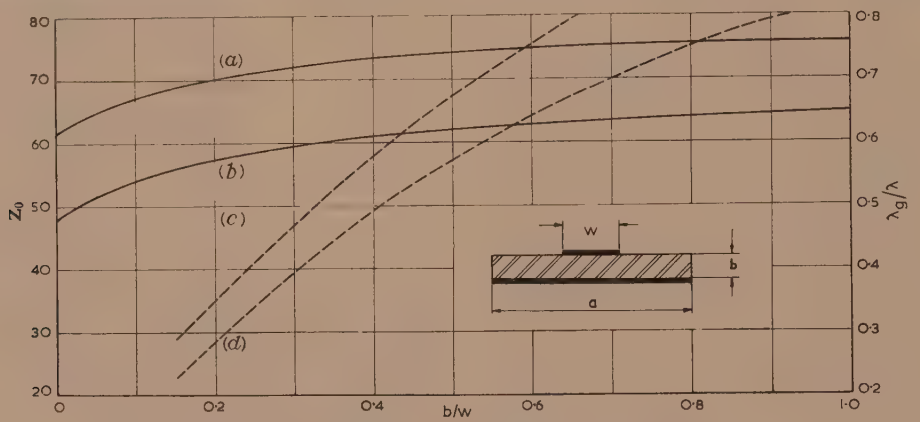


Fig. 7.—Characteristics of solid-dielectric microstrip line.

– – – – Characteristic impedance, Z_0.
——— Wavelength for $a > 3w$.
(*a*) and (*c*) P.T.F.E. Fibreglass, $\epsilon = 2{\cdot}65$.
(*b*) and (*d*) Silicone Fibreglass, $\epsilon = 4{\cdot}18$.

(3.2) Microstrip

The planar transmission line consisting of a strip conductor spaced from a ground plane by a solid dielectric layer is known as microstrip. Its general properties have been described by Grieg and Engelmann[35] and by Arditi,[1] who showed that microstrip is suitable as a wide-band means of transmitting microwave frequencies. The distance between the strip and ground-plane conductors is only a small fraction of a wavelength, but the composite dielectric means that a pure TEM-mode cannot be supported.

Several exact theories of propagation in microstrip have been outlined[20, 53, 54, 57] while an approximation based on the assumption of TEM-propagation has been given by Assadourian and Rimai.[4] If L and C are the inductance and capacitance per unit length, the characteristic impedance is given by

$$Z_0 = \sqrt{(L/C)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and the velocity of propagation by

$$v = 1/\sqrt{(LC)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Combining eqns. (11) and (12),

$$Z_0 = 1/Cv \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

which is the relation used by Arditi,[2] who measured C at low frequencies and v at microwave frequencies. The velocity of propagation is found to be constant over a wide range of frequencies, thus indicating that microstrip is not dispersive.

An accurate computation of the wavelength in microstrip has been made by Black and Higgins[10] for the case of an infinitely wide ground plane, values for two dielectric constants being given in Fig. 7 as a function of b/w. As the strip becomes wider, with b/w tending to zero, the wavelength approaches the value $\lambda/\sqrt{\epsilon}$. These results agree well with measurements made by Dukes[26] with the aid of an electrolytic tank for $a/w = 3$.

An analysis made by Bowness,[11] involving the use of a measured[6] constant for the fringing capacitance of the line, gave the characteristic impedance in ohms as

$$Z_0 = 10^4/3\epsilon^{1/2}[7 + 8{\cdot}83(w/b)] \quad . \quad . \quad . \quad . \quad (14)$$

This result assumes that the strip conductor is thin compared with its width and that $w \geqslant b$. A pseudo-TEM approach was adopted by Dukes[23] in computing the characteristic impedance of microstrip, the results obtained being in good agreement with measurements made at microwave frequencies. The values of ϵ assumed ranged from 1 to 10, and curves for 2·65 and 4·18 are given in Fig. 7.

The attenuation in microstrip is due to conductor, dielectric and radiation losses, and any measurement will give the sum of all three. For the dimensions generally used the conductor loss does not differ very greatly from that of an infinite parallel-plate system, given in decibels per metre by

$$\alpha_c = 27{\cdot}3R/\pi bZ' \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where R is the r.f. skin resistance in ohms per metre, and Z' is $377/\epsilon$ ohms, the wave impedance of the dielectric. The dielectric loss, given by eqn. (6), is usually higher than the conductor loss, while radiation loss is normally small.

Arditi[1] has shown that anomalous mode propagation causes the attenuation coefficient to vary near the launching point and measurements should be made at a distance of some 20 or more wavelengths away. Results for the attenuation and equiva-

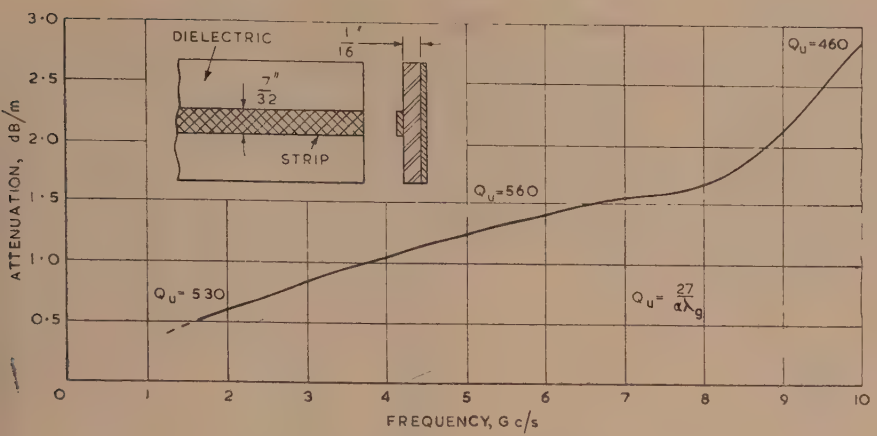


Fig. 8.—Attenuation of microstrip.
The dielectric material is Teflon-loaded Fibreglass.

lent unloaded Q-factor of a p.t.f.e.-impregnated Fibreglass line $\frac{1}{16}$ in thick are shown in Fig. 8. For dielectric $\frac{1}{8}$ in thick, the attenuation of 1·1 dB/m at 6 Gc/s is made up of about 0·3 dB/m for the conductor loss and 0·8 dB/m for dielectric loss: the unloaded Q-factor is about 700. When transmitting a continuous power of 300 watts at 3 Gc/s, a $\frac{1}{16}$ in line showed a rise in temperature of 50° C above the ambient of 20° C. Under pulse conditions, corona effects appeared at the edge of the strip conductor for a power of 15 kW at 9 Gc/s. If the dimensions of microstrip are scaled with the operating frequency, Dukes[23] shows that the total attenuation takes the form

$$\alpha = A + B\sqrt{f} \quad . \quad . \quad . \quad (16)$$

where A and B are constants depending on the materials.

(4) ELEMENTS AND COMPONENTS

(4.1) Elements

Theoretical formulae have been derived by Oliner[43, 44] for the equivalent circuit parameters of a variety of discontinuities in double-ground-plane lines. Such discontinuities possess purely reactive equivalent networks if they contain no dissipative elements and are symmetrical in shape and position. For example, a round hole in the strip conductor can be represented by a π-section of inductances, the values for a 75-ohm line being given in Fig. 9. Other cases considered included sharp corners and transverse posts, the computations being verified by measurements at 3 Gc/s.

The equivalent circuit of a step discontinuity in the strip conductor is an ideal transformer, the square of the turns ratio being given, to a first approximation, by the ratio of the characteristic impedances. This relation has been confirmed by Arditi[1] on experiments with microstrip at 4·7 Gc/s. The material was G–6 Fibreglass with strip width 26 mm, the width after the step ranging down to 1 mm. Owing to the low dispersion of microstrip, this type of discontinuity is not sensitive to frequency. This impedance-changing property enables quarter- and half-wavelength matching transformers to be constructed.

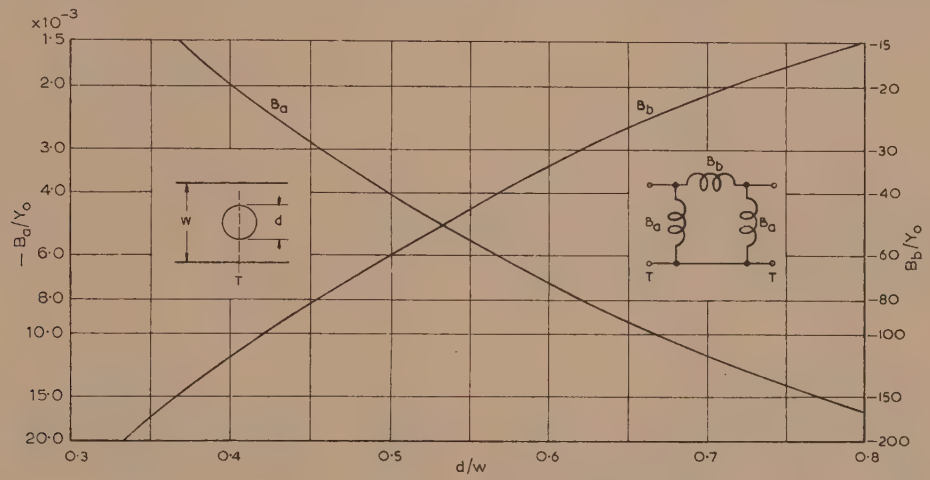


Fig. 9.—Effect of a round hole in a shielded strip-line.
The curves relate to a 75-ohm line.

Transverse posts tend to have a susceptance which varies with

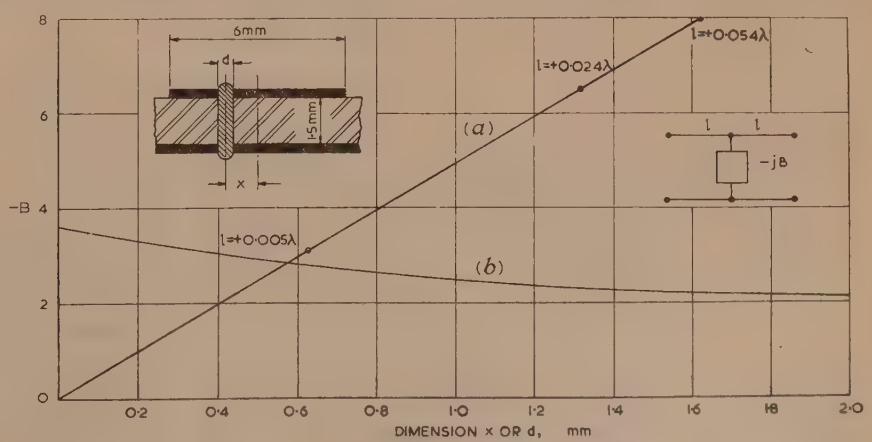


Fig. 10.—Shunt susceptance of a post in microstrip.
Frequency, 4·7 Gc/s; material, Fibreglass G-6. (*a*) B/d; (*b*) B/x with d = 0·7 mm.

frequency like a pure inductance, although radiation loss becomes important with large diameters. The variation of the equivalent normalized susceptance versus post diameter is shown for a microstrip line in Fig. 10(*a*), and the variation versus post position in Fig. 10(*b*). By spacing two posts at the proper distance Arditi and Elefant[3] obtained a resonant section of line, one example, using G–6 Fibreglass dielectric, giving at 5 Gc/s loaded and unloaded Q-factors of 23 and 165 respectively.

Stub sections in shunt with the main line behave like shunt reactances, which may be either inductive or capacitive and may have any value between zero and a maximum. This maximum is limited by line losses and depends on whether the end of the stub is open- or short-circuited. Double- or triple-stub sections may, like the other elements described, be used for impedance matching.[60]

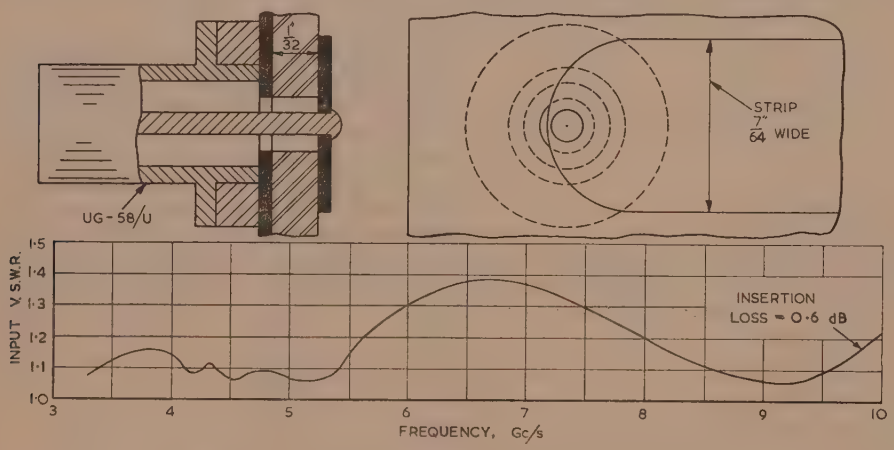


Fig. 11.—Wide-band coaxial-to-microstrip transducer.

(4.2) Transducers

Since planar transmission lines may exist in both the electrically balanced and unbalanced forms, some type of transition between them is occasionally required: three designs of such baluns have been described by Fubini,[32] the strip lines being all in one plane. Transducers between parallel-plate and coaxial lines are inherently broad band, with no lower frequency limit. The physical discontinuity between the two systems does, however, cause a reactance, the matching of which introduces some frequency sensitivity.

In a coaxial-to-triplate line transducer, Wild[60] used a series matching stub of width $0{\cdot}14\lambda$ and length $0{\cdot}12\lambda$ to give a v.s.w.r. better than 1·1 over a 10% band centred on 4·3 Gc/s. In a similar transition to a dielectric-sandwich line of characteristic impedance 50 ohms, Torgow and Griemsmann[57] obtained a v.s.w.r. better than 1·7 over the range 3–10 Gc/s. By removing the solid dielectric around the centre pin of a transducer from coaxial line to microstrip, as shown in Fig. 11, Arditi[1] extended the frequency range up to 10 Gc/s: the v.s.w.r. was better than 1·4 and the insertion loss lower than 0·6 dB.

At higher frequencies it is more convenient to change from parallel-plate line to waveguide. This may be achieved by a compact transition involving[23] a short length of coaxial line. Leakage of the order of −25 dB was found by Wild[60] in a direct transition from waveguide to triplate line: this was attributed to the simultaneous existence of both parallel-plate and coaxial transmission modes. The former was eliminated by short-circuiting bars which maintained the two outer plates at the same potential. In the transducer due to Arditi,[1] shown in Fig. 12, the normal rectangular section of the waveguide tapers gradually to a ridged guide which finally joins to the microstrip line.

Various bends, corners and twists may be constructed in parallel-plate lines. In microstrip, for example, a 45°-mitre or a smooth round corner produces no appreciable reflection, while the radiation loss at 5 Gc/s is less than 0·1 dB. A frequency insensitive match can be obtained by a length of tapered[50] stripline whose impedance at each end is equal to the impedance to be matched at that end.

(4.3) Components

Short-circuits in parallel-plate lines can be made to have a reflection coefficient more than 0·9 over a wide frequency band and, if necessary, they may be of variable position. Fixed and variable attenuators are made[60] in triplate line by inserting resistive card, or preferably metallized glass, between the strip conductor and one of the ground planes. A suitable surface resistivity is 100 ohms, and for maximum attenuation the resistive surface must be in contact with the strip. Zublin[62] found that lossy dielectric should preferably fill the space between the conductors, an air-gap of 0·015 in reducing the attenuation in a

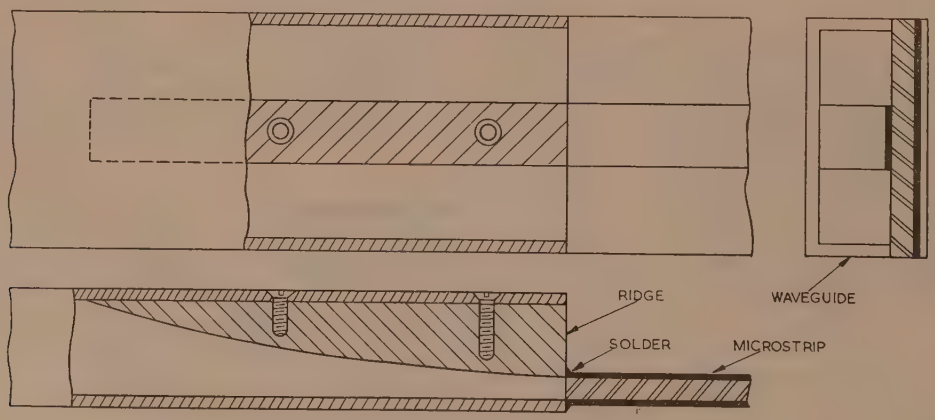


Fig. 12.—Waveguide-to-microstrip transducer.

particular case from 50 dB to 18 dB. Tapering the lossy card or medium improves the impedance match, and, on termination with a short-circuit, matched loads with a v.s.w.r. better than 1·1 over a wide frequency range have been obtained.

Cohn[16] has shown that the coupling between adjacent strip lines is weak and appreciable interchange of power is obtained only over a region of several wavelengths. Such a directional coupler has been constructed[65] for 500 Mc/s, while an example for 4·2 Gc/s, due to Wild,[60] based on a dielectric-sandwich line, is shown in Fig. 13(*a*). The attenuation depends critically on lead to inconveniently high values of characteristic impedance of the branch arms.

A more satisfactory type of directional coupler has been developed by Dukes[24] in which apertures in a common ground plane couple two back-to-back microstrip lines. According to the shape of the hole and the orientation of the two lines, either forward or reverse directivity is obtainable. Transverse slots have the advantage that excitation of the auxiliary guide is due only to the transverse component of magnetic field. The arrangement shown in Fig. 14 uses an array of such slots, the

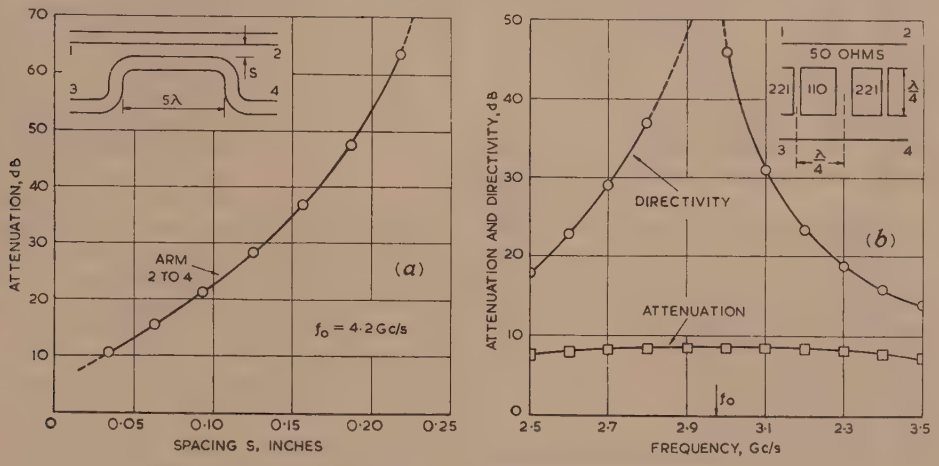


Fig. 13.—Types of triplate directional coupler.

(*a*) Parallel-line coupler.
(*b*) Branched-line coupler.

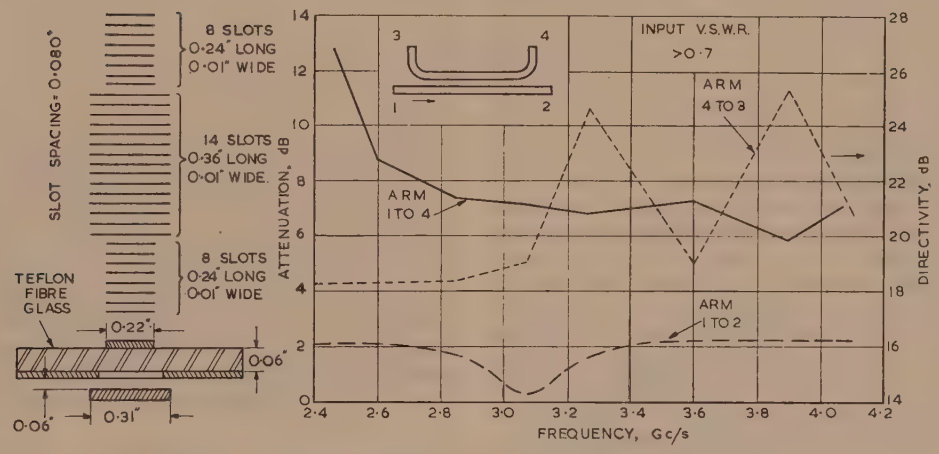


Fig. 14.—Phase-compensated back-to-back directional coupler.

he line spacing; moreover, long coupling regions tend to cause purious modes. Schiffman[67] has, however, used such coupled trips to provide a broad-band phase changer.

Fromm[29] obtained increased coupling with the branched-line rrangement shown in Fig. 13(*b*). This coupler is a binomial rray with quarter-wavelength spacing of the branch arms. The oupling of power is controlled by the length and impedance of he branches, while the directivity is a function of their spacing long the main line. The particular example shown is in high-Q-factor triplate line, the branch length and spacing being qual to 1·0 in, which is a quarter-wavelength at the design fre-uency of 2·96 Gc/s. The measured performance agreed well vith that computed. Branched line couplers are suitable only or narrow bandwidth, and for low coupling in microstrip they successive lengths being varied to give a taper to the coupling distribution in the interests of broad-band directivity. The coupling factor of a single slot is a simple linear function of frequency, but the array was made to give uniform coupling over a wide band by using the principle of phase compensation, in which the two transmission lines have unequal phase velocities. This condition is readily achieved in microstrip by omitting the solid dielectric in one of the lines.

The conventional hybrid-T, with series and shunt arms, cannot easily be realized in parallel-plane systems, and thus directional couplers and hybrid rings[1, 40, 48] are usually employed. Wild[60] has constructed a hybrid ring in triplate line which, over a band 4·1–4·5 Gc/s, had a v.s.w.r. better than 1·2, a coupling constant to 0·1 dB and a discrimination better than 40 dB. Wider band-

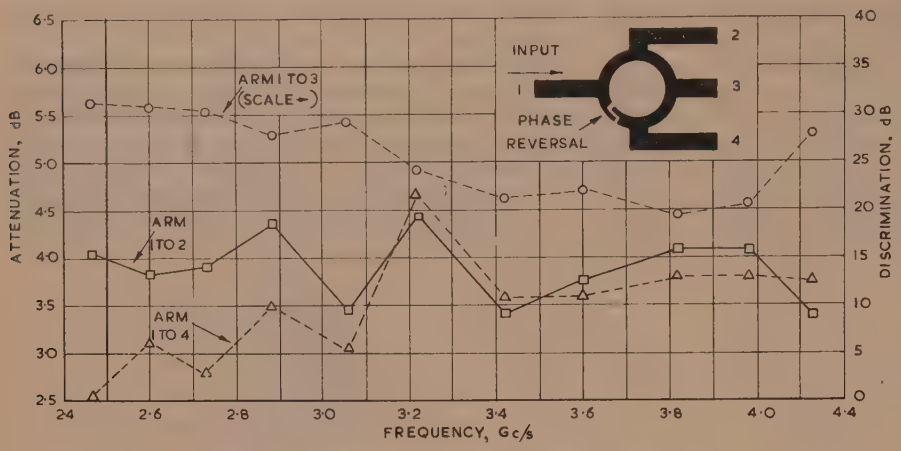


Fig. 15.—Reverse-phase hybrid ring in microstrip.

width was obtained by Dukes[23] by using 180° phase reversal in one part of the ring to equalize the path length: the electrical performance is shown in Fig. 15.

(5) CIRCUITS AND SYSTEMS

(5.1) Measurement Technique

The techniques used in determining such terminal-pair properties of parallel-plate components as impedance or attenuation may resemble those used with conventional transmission lines. Travelling-probe instruments are practicable in air-spaced triplate line, and Wild,[60] for example, replaced part of one of the ground planes by a plate containing a tapered slot through which a carriage-driven probe was inserted: the frequency range was 2–10 Gc/s. Cohn,[15] on the other hand, inserted the probe between the plates with the intention of obtaining greater accuracy for a given tolerance on strip flatness and alignment. At 1–5 Gc/s the measurement of impedance was believed to be correct within 1%. With either arrangement care is required to avoid leakage and spurious modes.

Slotted sections present difficulty in the case of microstrip, and it is preferable to use the Deschamps[19] technique and make measurements through a transducer with the aid of a coaxial line or waveguide instrument. The microstrip line is terminated in a short-circuit whose position is successively changed by amounts of $\lambda_g/8$. The corresponding reflection coefficients are plotted on a complex plane to form a circle. The chords joining opposite points intersect in a point that is related in a simple manner to the reflection and transmission coefficient of the transducer. The circle diagram gives the guide wavelength, and by repeating the measurement for two different line lengths the attenuation can be found. The calibrated transducer can now be used to measure the properties of the strip-line component under investigation.

(5.2) Circuit Devices

Wild[60] has constructed a ferrite phase changer in dielectric-sandwich line. A long ferrite rod was surrounded, first by polystyrene foam, and secondly, by a thin solid-dielectric sleeve. The whole was then encased by a metal cover to form a cylinder, the diameter being such that only TE_{11}-mode fields were excited. The two outside surfaces of the strip line were short-circuited to the metal cover, while the central strip was helically wound through several turns inside the dielectric sleeve. With this arrangement phase shift was accomplished through the influence of an axial magnetic field. To obtain a variable attenuator in microstrip, Arditi[1] placed a ferrite plate between the strip conductor and ground plane. The attenuation could be varied between zero and 12 dB by adjustment of the magnetic field: when operating as a phase-changer, relative shifts up to 150° were similarly obtained. More extensive measurements of the propagation coefficient of ferrite-filled microstrip have been made by Brodwin,[66] who found close agreement with analysis based on an infinite parallel-plate waveguide. If a piece of ferrite is placed on the edges of a parallel-plate line, gyromagnetic effects occur, and, using this principle, isolators have been developed by Fix[28] for triplate line and Lewin[41] for microstrip. When the r.f. magnetic fields of the strip line are in phase quadrature, a circularly polarized wave is produced in the fringing region. With a longitudinal static magnetic field, the structures exhibited non-reciprocal effects.

Semiconducting diodes are usually mounted in coaxial line or waveguide, and thus detectors and mixers in strip-line circuits must, at present, involve transducers. Carlson[13] used a single-ended mixer in microstrip, which, with directional-coupler feed for the backward-wave local oscillator, gave constant sensitivity over the frequency band 2–4 Gc/s. The balanced mixer of Dukes,[23] using a reverse-phase hybrid ring and reverse-polarity crystals, type CV2154 and CV2155, had an input v.s.w.r. of 2·0 and a decoupling of 20 dB over the range 2·8–4·4 Gc/s.

A flat gas-filled diode, approximately $\frac{1}{16}$ in thick and 3 in long, was mounted by Arditi[1] between the conductors of a microstrip line. Both hot- and cold-cathode valves, containing neon at a pressure of 1 mm Hg, were used. Variation of the steady discharge current from zero to 10 mA gave, at 5·6 Gc/s, a range of attenuation of 5–30 dB. The arrangement is suitable as a wide-band noise source, the effective noise factors being 24, 18, 15 and 13 dB at frequencies of 2, 4, 6 and 10 Gc/s respectively. As the discharge current is increased from zero the effective temperature rises and reaches a plateau at about 6 mA.

Assemblies of parallel-plate components have enabled equipment such as transmitter-receivers to be constructed. One example[23] for 3·8–4·2 Gc/s occupied two-thirds the volume of a similar unit constructed in waveguide. The equipment described by Fromm[29] for a similar frequency band contained two klystron local oscillators, signal and a.f.c. balanced mixers, hybrid ring, directional coupler, high-Q-factor strip-line resonator and several fixed and variable attenuators. The basic strip-line structure measured approximately 15 in × $6\frac{1}{2}$ in × 1 in, and, apart from the resonator, was manufactured as a single unit in triplate line.

(5.3) Frequency Filters

Frequency filters are usually based on the air-spaced triplate line to achieve the best possible Q-factor and to avoid leakage effects found with microstrip near the cut-off region. Dukes[25] has developed a series of low-pass re-entrant filters with sharp cut-off characteristics and total operating frequency ranges of several octaves. In one model, with a cut-off frequency of 4·1 Gc/s, the ground-plane spacing was $\frac{5}{16}$ in, the width of the strip conductor at the centre section was 0·017 in: the material used was 0·001 35 in copper foil on p.t.f.e.-Fibreglass supports. The individual sections of the filter were given tapered parameters to provide sharp cut-off and a good impedance match. The transmission-line filters of Fromm[29] employed series and shunt elements; examples included a high-pass design for 600 Mc/s and low-pass for 1·5 Gc/s and 3·0 Gc/s.

Direct-coupled cavity filters, using end-supported strips, were used by Moore and Michelson[42] for 6·725 Gc/s. The width of

the strip was $\frac{1}{2}$ in, and with a $5\lambda/4$ length the Q-factor was 2 100. Tuning was accomplished by changing the resonator length, the end support being provided with folded half-wave chokes to facilitate this movement. Frequency drift induced by humidity was eliminated by surrounding the resonator with polystyrene foam, and coupling to the resonator was by means of another strip separated by a small gap. Four such resonators, synchronously tuned and coupled to give a maximally-flat filter, possessed a half-power bandwidth of 30 Mc/s and an insertion loss in the pass band of 3·5 dB. In the filters of White and Bradley[59] the series-capacitance coupling between the strip-line resonators was realized by making a dovetail-shaped gap in the centre conductor. A six-stage maximally-flat filter had a 10% bandwidth with minimum insertion loss and 40 dB rejection at 12% from the centre frequency of 3·35 Gc/s.

The frequency filter of Cohn and Coale,[17] shown in Fig. 16(*a*), possesses directional properties. Energy entering arm 1 excites a wave travelling clockwise in the loop, the parallel strips having reverse coupling. The second directional coupler removes energy from the loop through arm 4 of the network. Resonance occurs at the frequency at which the loop is one wavelength in perimeter, and it can be shown that all the input power emerges from arm 4 at this frequency, while at frequencies sufficiently removed from it virtually all the power emerges from arm 2. If the filter is properly adjusted, no power emerges from arm 3 at any frequency, nor is any reflected from arm 1. Greater selectivity may be achieved by the use of two or more loops coupled to each other in cascade as in Fig. 16(*b*). The circuits of Figs. 16(*c*) and (*d*) contain two separate strip lines at points separated by an odd number of quarter wavelengths: the resonant strips are made dissimilar, so that the coupling points on one are at opposite polarities and on the other at the same polarity. The resonant strips of Fig. 16(*e*) are identical, but an effective polarity reversal is achieved through the use of a separation of the coupling points one-half wavelength greater on one line than on the other. Greater selectivity is achieved in Fig. 16(*f*) by using resonators in cascade. An experimental model of the circuit of Fig. 16(*e*) gave, at the design frequency of 884 Mc/s and with input in arm 1, a v.s.w.r. of 1·24 and attenuations to arms 2 and 4 of 18 and 0·8 dB respectively: these results verified the performance of the circuit as a directional filter.

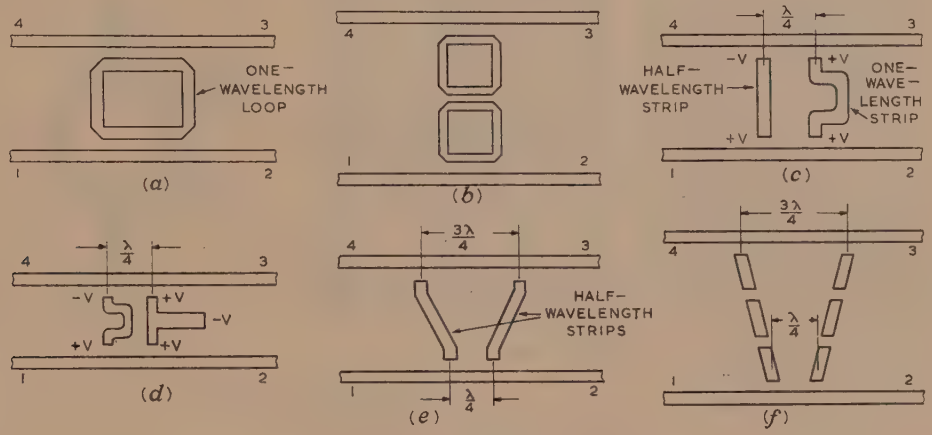


Fig. 16.—Directional filters in strip line.

(5.4) Antenna Circuits

Parallel-plate lines are suitable[31] for feeding and exciting antenna elements. Dahlman[18] describes a multiple-feed system in solid-dielectric sandwich line which involves 50/25-ohm tapers, right-angle bends, T-junctions and transducer to coaxial line. The dielectric was 2 mm-thick polystyrene, and the strip was 0·04 mm-thick copper foil, the whole system being matched at 3 Gc/s and having an attenuation of 0·3 dB. Power-split systems in strip line enable the distribution along an antenna array to be controlled: the split ratio depends upon the widths of the strip and is not limited to exponents of 2.

A typical multiple-antenna feed, designed by Sommers,[55] is shown in Fig. 17(*a*). The input line to the power splitter has a characteristic impedance of 22·5 ohms and feeds, through a 3 : 1 power division, two lines with characteristic impedances of 30 and 90 ohms. The 30-ohm line divides, in turn, into a 45-ohm line and a second 90-ohm line, giving in this case a 2 : 1 power split. Finally, the power in the 45-ohm line divides equally into two more 90-ohm lines. Measurements on such a progressive power divider showed a maximum deviation of only 0·6 dB over a 10% frequency band centred on 9·375 Gc/s.

Parallel-plate techniques have been found suitable[33, 34] for antennae of the surface type. Radiating slots cut in the ground plane provide antennae free of projections, thus enabling flush mounting to be adopted on, for example, high-speed aircraft.[56] The polar pattern radiated by the surface wave supported by a metal-backed solid-dielectric sheet can be modified and controlled by forming metal strips on the outside face of the dielectric. Fubini[32] has shown how a conventional Franklin antenna can be realized at microwave frequencies by a collinear array of half-wave strip elements separated by folded half-wave sections of line.

Sommers[55] used a slot cut transversely in one of the ground planes of a solid-dielectric sandwich line as a radiating element. Lateral leakage, due to the slot, from between the outer plates was eliminated by inserting short-circuiting pins to suppress unwanted modes. The element was $0{\cdot}6\lambda$ long and $0{\cdot}05\lambda$ wide in series with a 90-ohm triplate line which terminated in an open-circuit at a distance $0{\cdot}1\lambda_g$ from the centre of the slot. The final antenna, consisting of a 4×4 slot array, was arranged as in Fig. 17(*a*): the H- and E-plane radiation patterns are shown in Figs. 17(*b*) and (*c*) respectively. The H-plane array, with $1{\cdot}92\lambda$ spacing between the centres of the outside slots, has a beamwidth of 20° at half power, while the E-plane array, with $1{\cdot}8\lambda$ spacing between slot centres, has a 23·5° beamwidth. The high side-lobe intensities were due primarily to mutual coupling between diagonal slots.

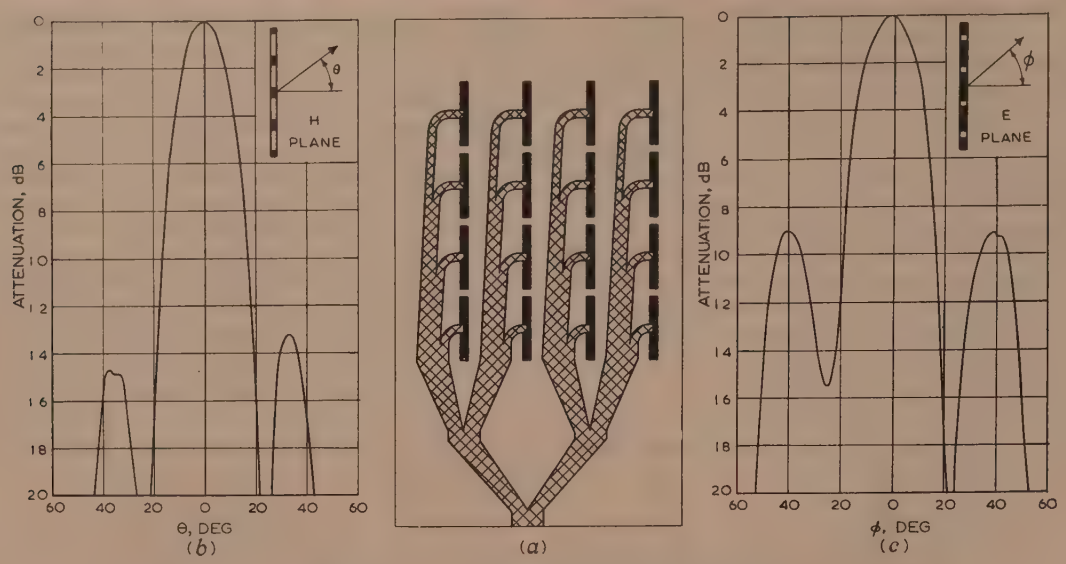


Fig. 17.—Slot antenna with triplate construction.

(*a*) Feed arrangement of the 4 × 4 slot array.
(*b*) and (*c*) *H*- and *E*-plane radiation patterns. The frequency is 9·375 Gc/s.

(6) MATERIALS AND MANUFACTURE

(6.1) Materials and Properties

The basic material for the unsymmetrical solid-dielectric planar line consists of a board of low-loss solid dielectric coated on both sides with metal foil. The strip-line circuit is made on one side of this sheet. With a symmetrical line a further sheet, clad only on one side, is laid on the first so that the strip circuit is centred between the two ground planes.

For laboratory and experimental purposes, metal foil can be bonded to thin sheets of dielectrics such as polythene and polystyrene. A weak bond of copper foil may be achieved with controlled application of heat and pressure, but Bowness[11] used Durofix and polystyrene cement mixed in suitable proportions with polythene and polystyrene solvents. With care, air pockets can be eliminated so that the foil has a smooth surface. The metallic surface may, alternatively, be applied by means of conducting silver paint or ink made in grades specially intended for application to plastics. Such paints enable, for example, strips 0·1 in wide to have an r.f. resistance less than 0·1 ohm/in.

The improved mechanical properties necessary for a material intended for production have been achieved in commercially available metalclad boards using p.t.f.e. or materials reinforced with Fibreglass cloth. The latter class includes a variety of boards, and Ringenbach and Cooper[51] have made measurements of phase velocity and attenuation at 1 Gc/s with impregnation by phenolic, epoxy and polyester resins, and p.t.f.e. Under typical conditions the attenuations in decibels per wavelength were 0·14, 0·08, 0·028 and 0·016 respectively.

The material used most extensively is p.t.f.e.-impregnated Fibreglass, although silicone-impregnated Fibreglass, with intermediate cost and performance, has found limited application. Impregnated Fibreglass boards tend to vary in composition, give the metal foil a slightly corrugated surface, and are not resistant to prolonged exposure in humid conditions. The available sheet sizes are about 36 in × 16 in with thicknesses from 0·002 in to $\frac{5}{8}$ in. Typical thickness tolerances are 0·007 5 in on $\frac{1}{16}$ in and 0·012 in on $\frac{1}{8}$ in. The thickness of the copper foil is usually 0·001 35 in or 0·002 7 in and the bond strength is 7–9 lb per inch of foil. The electric strength is about 350 volts per mil, and the materials will operate continuously at temperatures up to 200° C. Some other important properties of these materials are given in Table 1.

Table 1

PROPERTIES OF DIELECTRIC SHEET MATERIALS

Property	Pure Teflon	Teflon Fibreglass	Silicone Fibreglass
Weight, lb/in³	0·078	0·078	0·061
Dielectric constant at 1 Mc/s	1·9	2·65	3·9
Loss tangent at 1 Mc/s ..	0·000 55	0·000 7	0·001 5
Water absorption, % ..	0·07	0·03	0·30

The silicone-impregnated Fibreglass is manufactured by coating the glass fabric with a synthetic silicone resin which has first been dissolved. Drying ovens remove the solvent, and the impregnated sheets are cut to size and piled to the required thickness. Metal foils are cemented on to the outer laminates by the use of an adhesive film, and heat and pressure are applied. The resin flows and cures into a homogeneous material resulting in a finished product that is infusible and insoluble. The Teflon product is produced in a similar manner, except that the Teflon in dispersion is applied to the Fibreglass cloth by a dipping process, then dried and fused at elevated temperatures. The resultant product, as before, is laminated into sheets of various thicknesses.

To avoid the time and trouble of constructing individual components, a microstrip kit has been made available commercially.[63] It contains an assortment of circuit components, connectors to join them together electrically and an easel to which they may be fastened mechanically. Coaxial cables may be connected to fittings mounted on certain components, and the circuits may be used over the frequency range 1–8 Gc/s. This kit provides a simple and rapid means of building microwave assemblies.

(6.2) Printing Technique

For experimental purposes the strip-line circuit may be cut out with a sharp knife: alternatively the board can be covered with transparent adhesive cellulose tape which is cut along marking lines and the surplus is peeled off. The unwanted copper foil

is then etched away leaving the circuit in position. Planar strip circuits can, however, be made economically by the printing techniques which are now well established[27] in the radio industry. A suitable etching process is based on the 'cold-top' enamel method described, for example, by Bittner[9] and others.[64] The process begins with the engineer making an exact-size design of the microwave circuit to be etched. A dimensionally accurate ink drawing is then prepared, using as large a scaling factor as possible to reduce drawing errors. For example, an error of $\pm 0{\cdot}015$ in will reduce to $\pm 0{\cdot}000\,75$ in using a 20 : 1 reduction ratio. The drawing material may be tracing cloth, but frosted acetate is better, since it has higher dimensional stability and does not allow the ink to feather. A typical drawing for a hybrid ring is shown in Fig. 18(*a*).

The master drawing is then reproduced as a photographic negative of the desired size. If necessary the image is measured on an optical comparator to determine the accuracy. It is important that the base for the negative be insensitive to any processing or environmental conditions that it might experience during the repetitive exposures of quantity manufacture. Most of the high-contrast emulsions on glass are suitable, together with films using a stable base material such as ortho-vinyl. Even a good film would show a shrinkage factor of 1·005, and this must be allowed for in taking the photograph.

These negatives may be pieced together, either to form more complex circuits or to provide multiple images for even greater economy in production. A negative containing four images is shown in Fig. 18(*b*), but up to 30 or more can often be used in practice, depending on the size of plastic sheet available and the capacity of the processing equipment.

The next stage is to clean all dirt and oxides off the copper laminations and to apply the photo-sensitive enamel to one side of the sheet. To obtain uniformity the enamel may be flowed on carefully, the process being carried out in ordinary interior illumination. The sensitized sheet is dried under infra-red lamps, being whirled rapidly for a controlled period.

The sensitized copper laminate is then placed under the negative, both being sandwiched in a vacuum printing frame to provide close contact and thus eliminate errors due to parallax. The frame is exposed for several minutes to an intense ultra-violet light source such as an arc lamp. This light hardens the enamel in the places to which it has access through the negative.

The sheet is placed in a developer which washes away the soft coating. The developer may consist of a mixture of dye and alcohol, the dye showing whether the process is proceeding correctly. The developed sheet is next washed and dried. Any defects in the design are touched up with acid-resisting paint, and the copper laminate on the reverse side is entirely covered with a clear acid-resistant plastic coating that is retained permanently.

The unprotected copper is etched away in a bath of ferric perchloride. This process may be temperature-controlled with motor-driven paddles to mix the solution to give a uniform action. The sheet is washed, and as a final step the components are given a thin plating of silver. After drying, the finished article is obtained as in Fig. 18(*c*), the sheet being divided, if necessary, to give the individual circuits. There is a small enlarging action during the etching process, and, if this is allowed for, dimensions can be maintained to $\pm 0{\cdot}001$ in.

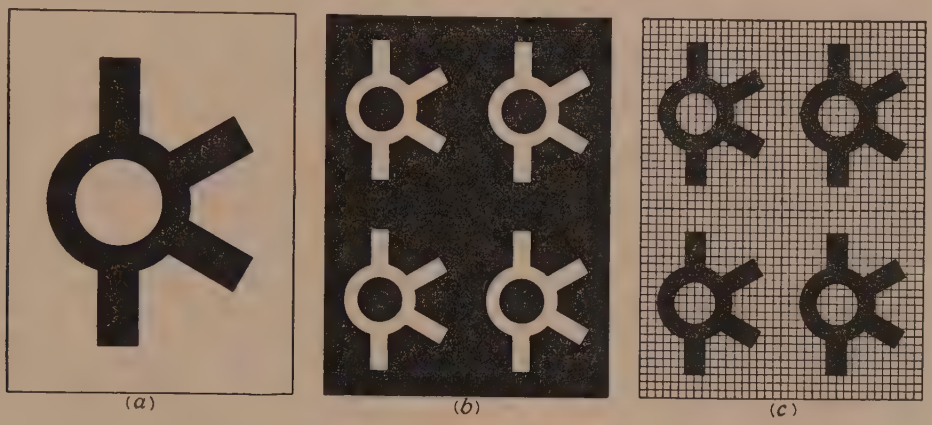


Fig. 18.—Stages in the printing of a microwave circuit.
(*a*) Master drawing.
(*b*) Negative with multiple circuits.
(*c*) Finished components.

(7) ACKNOWLEDGMENTS

The author is indebted to the Institute of Radio Engineers for permission to reproduce several illustrations from their journals and to Mr. J. M. C. Dukes for much useful discussion on this subject and for reading the manuscript.

(8) BIBLIOGRAPHY

(1) Arditi, M.: 'Characteristics and Applications of Microstrip for Microwave Wiring', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 31.
(2) Arditi, M.: 'Experimental Determination of the Properties of Microstrip Components', *Electrical Communication*, 1953, **30**, p. 283; *Convention Record, Institute of Radio Engineers*, 1953, Part 10, p. 27.
(3) Arditi, M., and Elefant, J.: 'Microstrip applied to Band-Pass Microwave Filters', *Electrical Communication*, 1955, **32**, p. 52.
(4) Assadourian, F., and Rimai, E.: 'Simplified Theory of Microstrip Transmission Systems', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1651.
(5) Barrett, R. M.: 'Etched Sheets serve as Microwave Components', *Electronics*, 1952, **25**, p. 114.
(6) Barrett, R. M.: 'Microwave Printed Circuits—a Historical Survey', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 1.
(7) Barrett, R. M., and Barnes, M. H.: 'Microwave Printed Circuits', *Radio and T.V. News*, 1951, **46**, p. 16.
(8) Begovitch, N. A.: 'Capacity and Characteristic Impedance of Strip Transmission Line with Regular Inner Conductors', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 127.
(9) Bittner, B. J.: 'Photoetching Techniques for the Precision Fabrication of Waveguide Antenna Arrays', *Symposium on Microwave Printed Circuits, Tufts College*, October, 1954.
(10) Black, K. G., and Higgins, T.: 'Rigorous Determination of the Parameters of Microstrip Lines', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 85.
(11) Bowness, C.: 'Strip Transmission Lines', *Electronic Engineering*, 1956, **28**, p. 2.

(12) BRADLEY, E. H.: 'Design and Development of Strip-Line Filters', *Transactions of the Institute of Radio Engineers*, 1956, **MTT-4**, No. 2, p. 86.
(13) CARLSON, E.: 'A Broadband Microstrip Crystal Mixer with Integral D.C. Return', *ibid.*, 1955, **MTT-3**, No. 2, p. 175.
(14) COHN, S. B.: 'Characteristic Impedance of the Shielded-Strip Transmission Line', *ibid.*, 1954, **MTT-2**, p. 52.
(15) COHN, S. B.: 'Problems in Strip Transmission Lines', *ibid.*, 1955, **MTT-3**, No. 2, p. 119.
(16) COHN, S. B.: 'Shielded Coupled-Strip Transmission Lines', *ibid.*, 1955, **MTT-3**, No. 5, p. 29.
(17) COHN, S. B., and COALE, F. S.: 'Directional Channel-Separation Filters', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 1018.
(18) DAHLMAN, B. A.: 'A Double-Ground-Plane Strip-Line System for Microwaves', *Proceedings I.E.E.*, Paper No. 1852 R, July, 1955 (**102**, p. 488), and *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 5, p. 52.
(19) DESCHAMPS, G. A.: 'New Chart for the Solution of Transmission-Line and Polarization Problems', *Electrical Communication*, 1953, **30**, p. 247, and *Transactions of the Institute of Radio Engineers*, 1953, **MTT-1**, p. 5.
(20) DESCHAMPS, G.: 'Theoretical Aspects of Microstrip Waveguides', *Transactions of the Institute of Radio Engineers*, 1954, **MTT-2**, No. 1, p. 100.
(21) DUKES, J. M. C.: 'Transmission-Line Termination', *Wireless Engineer*, 1955, **32**, p. 266.
(22) DUKES, J. M. C.: 'Characteristic Impedance of Air-Spaced Strip Transmission Line', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 876.
(23) DUKES, J. M. C.: 'The Application of Printed-Circuit Techniques to the Design of Microwave Components', *Proceedings I.E.E.*, Paper No. 2401 R, August, 1957 (**105 B**, p. 155).
(24) DUKES, J. M. C.: 'Broad-Band Slot-Coupled Microstrip Directional Couplers', *ibid.*, Paper No. 2402 R, August, 1957 (**105 B**, p. 147).
(25) DUKES, J. M. C.: 'Re-Entrant Transmission Line Filter using Printed Conductors', *ibid.*, Paper No. 2444 R, November, 1957 (**105B**, p. 173).
(26) DUKES, J. M. C.: 'An Investigation into Some Fundamental Properties of Strip Transmission Lines with the Aid of an Electrolytic Tank', *ibid.*, Paper No. 1991 R, May, 1956 (**103 B**, p. 319).
(27) DUMMER, G. W. A., and JOHNSTON, D. L.: 'Printed and Potted Electronic Circuits', *ibid.*, Paper No. 1407 R, November, 1952 (**100**, Part III, p. 178).
(28) FIX, O. A.: 'A Balanced Strip-Line Isolator', *Convention Record of the Institute of Radio Engineers*, 1956, Part 5, p. 99.
(29) FROMM, W. E.: 'Characteristics and Some Applications of Strip-Line Components', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 13.
(30) FROST, A. D., and MINGINS, C. R.: 'Microwave Strip Research at Tufts College', *ibid.*, 1955, **MTT-3**, No. 2, p. 10.
(31) FROST, A. D., MCGEOCH, C. R., and MINGINS, C. R.: 'The Excitation of Surface Waveguides and Radiating Slots by Strip-Circuit Transmission Lines', *ibid.*, 1956, **MTT-4**, No. 4, p. 218.
(32) FUBINI, E. G.: 'Strip-Line Radiators', *ibid.*, 1955, **MTT-3**, No. 2, p. 149.
(33) FUBINI, E. G., FROMM, W., and KEEN, H.: 'New Techniques for High-Q Strip Microwave Components', *Convention Record of the Institute of Radio Engineers*, 1954, Part 8, p. 91.
(34) FUBINI, E., FROMM, W., and KEEN, H.: 'Microwave Applications of High-Q Strip Components', *ibid.*, 1954, Part 8, p. 98.
(35) GRIEG, D. D., and ENGELMANN, H. F.: 'Microstrip—A New Transmission Technique for the Kilomegacycle Range', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1644, and *Proceedings of the Institute of Radio Engineers, Australia*, 1954, **15**, p. 13.
(36) HAYT, W. H.: 'The Mutual and Input Impedance of Dipole Strips between Parallel Planes', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 114.
(37) HAYT, W. H.: 'Potential Solution of a Homogeneous Strip-Line of Finite Width', *ibid.*, 1955, **MTT-3**, No. 4, p. 16.
(38) HORGAN, J. D.: 'Coupled Strip Transmission Lines with Rectangular Inner Conductors', *ibid.*, 1957, **MTT-5**, No. 2, p. 92.
(39) JONES, E. M. T., and BOLLJAHN, J.: 'Coupled Strip-Line Filters and Directional Couplers', *ibid.*, 1956, **MTT-4**, p. 75.
(40) KOSTRIZA, J. A.: 'Microstrip Components', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1658.
(41) LEWIN, L.: 'A Resonance Absorption Isolator in Microstrip for 4 Gc/s', *Proceedings I.E.E.*, 1956, **104 B**, Supplement No. 6, p. 364.
(42) MOORE, J. F., and MICHELSON, M.: 'Resonator and Preselector in Balanced Strip-Line', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 170.
(43) OLINER, A. A.: 'The Radiation Conductance of a Series Slot in Strip Transmission Line', *Convention Record of the Institute of Radio Engineers*, 1954, Part 8, p. 78.
(44) OLINER, A. A.: 'Equivalent Circuits for Discontinuities in Balanced Strip Transmission Line', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 134.
(45) OLINER, A. A.: 'Theoretical Developments in Symmetrical Strip Transmission Line', *Proceedings of the Symposium on Modern Advances in Microwave Techniques, Polytechnic Institute of Brooklyn*, November, 1954, p. 379.
(46) PACKARD, K. S.: 'Machine Methods make Strip Transmission Lines', *Electronics*, 1954, **27**, p. 148.
(47) PARK, D.: 'Planar Transmission Lines, I and II', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 3, p. 8, and No. 5, p. 7.
(48) PASCALAR, H. G.: 'Strip Line Hybrid Junction', *ibid.*, 1957, **MTT-5**, No. 1, p. 23.
(49) PEASE, R. L., and MINGINS, C. R.: 'Universal Approximate Formula for Characteristic Impedance of Strip Transmission Line with Rectangular Inner Conductor', *ibid.*, 1955, **MTT-3**, No. 2, p. 144.
(50) PRIMOZICH, F. G., SCHATZ, E. R., and WOODFORD, J. B.: 'A Tapered Strip Line for Pulse Transformer Service', *Electrical Engineering*, 1955, **74**, p. 908, and *Transactions of the American I.E.E.*, 1955, Part 1, p. 158.
(51) RINGENBACH, M. E., and COOPER, H. WARREN: 'Measurement of Attenuation and Phase Velocity of Various Laminate Materials at L-Band', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 87.
(52) SCIEGIENNY, J.: 'Strip above Ground Plane Transmission System', *Quarterly Progress Report of the Research Laboratory of Electronics M.I.T.*, 1953, Part I, p. 53.
(53) SCIEGIENNY, J., and SCHETZEN, M.: 'Strip Transmission Systems', and 'Theoretical Analysis of a Strip Transmission System', *ibid.*, 1953, Part XI, p. 83.
(54) SCHETZEN, M.: 'Printed Microwave Systems', *Technical Report* 289, *Research Laboratory of Electronics, M.I.T.*, September, 1954.
(55) SOMMERS, D. J.: 'Slot Array Employing Photoetched Tri-plate Transmission Line', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 157.
(56) SOMMERS, D. J.: 'Photoetched Antennae for Supersonic Aircraft', *Electronics*, 1955, **28**, p. 130.
(57) TORGOW, E. N., and GRIEMSMANN, J. W. E.: 'Miniature Strip Transmission Line for Microwave Applications', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 57.
(58) WHITE, D. R. J.: 'Developments in Printed Microwave Components', *Electronic Industries and Tele-Tech*, 1957, **16**, No. 11, p. 63.
(59) WHITE, D. R. J., and BRADLEY, E. H.: 'Band-Pass Filters using Strip-Line Techniques', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 163, and *Electronics*, May, 1955, p. 152.
(60) WILD, N. R.: 'Photoetched Microwave Transmission Lines', *Tele-Tech*, February, 1955, p. 68, and *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 21.
(61) WU, T. T.: 'Theory of the Microstrip', *Journal of Applied Physics*, 1957, **28**, p. 299.
(62) ZUBLIN, K. E.: 'Strip Type Components for 2 000 Mc/s Receiver Head End', *Transactions of the Institute of Radio Engineers*, 1955, **MTT-3**, No. 2, p. 65.
(63) 'Microstrip Kit', *Electrical Communication*, 1954, **31**, p. 214.
(64) 'Manufacture of Microstrip', *ibid.*, 1952, **29**, p. 250.
(65) 'Printed Circuit Directional Coupler', *Electronic and Radio Engineer*, 1957, **34**, p. 133.
(66) BRODWIN, M. E.: 'Propagation in Ferrite-Filled Microstrip', *Transactions of the Institute of Radio Engineers*, 1958, **MTT-6**, No. 2, p. 150.
(67) SCHIFFMAN, B. M.: 'A New Class of Broad-Band Microwave 90-Degree Phase Shifters', *ibid.*, 1958, **MTT-6**, No. 2, p. 232.

621.371.029.6

The Institution of Electrical Engineers
Paper No. 2779 E
Mar. 1959
©

OPTICAL TECHNIQUES AT MICROWAVE FREQUENCIES

By A. F. HARVEY, D.Phil., B.Sc.(Eng.), Member.

(*The paper was first received 23rd April, and in revised form 19th September*, 1958.)

SUMMARY

The paper reviews the application of optical principles and techniques at microwave frequencies. A brief discussion of radiation and diffraction in the far and near fields is first given. This is followed by an account of the various types of artificial dielectric. Methods for reducing and enhancing surface reflectivity are examined, and several instruments and components, which are especially useful at millimetre wavelengths, are considered. The survey concludes with a bibliography.

(1) RADIATION AND DIFFRACTION

(1.1) Diffraction Theory

Optical techniques can be applied at microwave frequencies, provided account is taken of the difference in wavelength. The dimensions used for radiating surfaces and apertures are usually comparable with the free-space wavelength λ: thus, whereas an aperture width of 10 cm represents about 180 000 wavelength of green light ($\lambda = 5\,500 \times 10^{-8}$ cm), it is only 10 wavelengths for typical microwave radiation ($\lambda = 1$ cm). Hence the study and application of microwave radiation must include the principles of diffraction: that is, the study of the effect of the phase and amplitude distribution across radiating apertures. The application of optical principles to microwave problems has a long history,[19, 21, 81, 119] but has only become practicable with the development of stable coherent sources of power, accurate measuring instruments, improved dielectrics and methods of control of surface phenomena.

Radiation in free space takes place when there are oscillating distributions of current or voltage along a conductor or across an aperture. These oscillations set up electric and magnetic fields in the region around the radiator which represent both stored and radiated energy, while there is also energy dissipated in the conducting surfaces. The radiated energy determines the input resistance of the radiator, and associated with it are electric and magnetic fields at distant points. The directional characteristics of this radiated energy are described by the radiation pattern or polar diagram. This is a graph of the electric field intensity at a fixed distance from the source as a function of angle measured in the particular plane for which the pattern is drawn. The gain of a radiator in a given direction is defined as the ratio of the maximum radiated power per unit solid angle to the radiated power per unit solid angle from an isotropic radiator, assuming the same total radiated power in each case.

The simplest radiators are electric and magnetic dipoles. The former consists of a small conductor along which an r.f. current flows, while the latter consists of a small loop in which r.f. current flows. The field distributon of an electric dipole is shown in Fig. 1(*a*), while Fig. 1(*b*) gives the polar diagram in the electric, or *E*-plane, containing the electric flux lines. The corresponding diagram in the magnetic or *H*-plane, containing the magnetic flux lines, is given in Fig. 1(*c*). For a magnetic loop lying in the *xy*-plane the polar diagrams represent the magnetic field. A common microwave radiator is the centre-fed half-wavelength dipole, which may also be regarded as a resonant transmission line. The polar diagram in the *E*-plane is given[143] by

$$E_\theta = \cos\left(\frac{\pi}{2}\cos\theta\right)/\sin\theta \quad . \quad . \quad . \quad . \quad (1)$$

and the gain is 1·65. More complicated radiators include arrays of elementary sources and radiating surfaces and apertures.

The general problem of calculating the electromagnetic field at a distance in terms of the current distribution at the aperture can be formulated, sometimes precisely and usually to a very good approximation. The exact methods tend to be difficult to apply and some approximation is required for general use. One approximation, which is often used, is that of simple geometric ray optics, in which the microwave energy is assumed to follow paths determined by the ordinary laws of reflection and refraction. Possible ray paths are given by Fermat's principle[161] as the curves along which the path length is stationary with respect to infinitesimal variations in path. It follows that the rays in a homogeneous medium are straight lines. Optical-ray methods

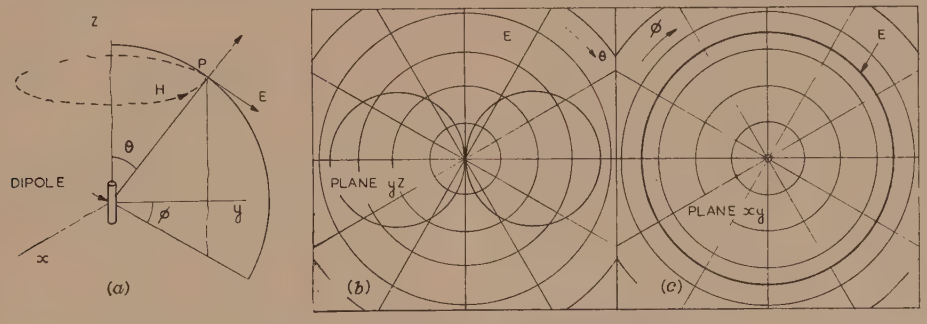


Fig. 1.—Radiation field of an electric dipole.

(*a*) Directions of the electric and magnetic fields.
(*b*) Electric field for the electric plane.
(*c*) Electric field for the magnetic plane.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

This is an 'integrating' paper. Members are invited to submit papers in this category, giving the full perspective of the developments leading to the present practice in a particular part of one of the branches of electrical science.

Dr. Harvey is at the Royal Radar Establishment.

can also be used to estimate radiation fields which are continuous functions of angle. The essential requirement is that there should be a one-to-one correspondence between each point on the aperture and each direction in the radiation diagram. These optical methods assume that the aperture is very large compared with the wavelength and thus the results are not accurate. The simplicity of these methods often outweighs the crudeness of their approximation to the required diagram.

Another method of analysis is that of Huyghens, in which[11] each point of a wavefront is assumed to be the source of a wavelet that travels out in an expanding sphere. At any later instant the new wavefront is given by the envelope of the secondary waves. Unlike the dipole, the Huyghens source considered as an elementary radiator takes no account of polarization and gives only the scalar field intensity: it is thus only an approximation for electromagnetic waves.

The scalar form of Huyghens's principle was extended by Kirchhoff, whose theory enables[11] the field at points distant from a radiating aperture to be calculated in terms of the current distribution with sufficiently good accuracy for most cases. If the amplitude and phase of this distribution are known, it is merely necessary to add vectorially the contributions from the various current elements. Thus, consider the radiating surface which lays down a field at a distant point P. If E is the electric field over a small area dS of the radiating surface S, E_p is the field at P due to dS, and $\partial/\partial n$ denotes differentiation along the normal drawn outward to the surface S, then the Huyghens–Kirchhoff formula gives

$$E_p = \frac{1}{4\pi}\left\{E\frac{\partial}{\partial n}\left[\frac{1}{r}\exp(-j\beta r)\right] - \frac{1}{r}\exp(-j\beta r)\frac{\partial E}{\partial n}\right\}dS \quad . \quad (2)$$

where β is the propagation coefficient, $2\pi/\lambda$. The total field is obtained by integrating this over the aperture. Thus for a uniformly illuminated rectangular aperture of width a the complex polar diagram in the far field, with r large, is

$$\mathrm{E}(\theta) = \frac{1}{\lambda r}\int_{-a/2}^{a/2} A\exp(j\beta y\sin\theta)dy \quad . \quad . \quad . \quad . \quad (3)$$

where y is the distance from the aperture centre, θ is the angle from the aperture normal and A is the aperture distribution of current. This equation refers to that plane of the rectangular aperture which contains the width a along the y-axis and the angle θ, and is based on the assumption that the variables can be separated for the other plane also.

Another approximate estimation of the radiation field makes use of the fact that the field at all points in front of a plane aperture, both in the far- and near-field regions, may be regarded as arising from the interference of plane waves travelling in various directions. The amplitude and phase of the waves, as a function of their direction of travel, constitute an angular spectrum which, appropriately expressed, is the Fourier transform of the aperture distribution. This was first pointed out by Stratton[143] and was critically examined by Booker and Clemmow,[20] who extended the theory.

It is well known that a fluctuation $\mathrm{f}(t)$ can be analysed into a series of oscillations of different frequencies $\mathrm{S}(\omega)$ by an expression of the form

$$\mathrm{f}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{S}(\omega)\,\varepsilon^{j\omega t}d\omega \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Furthermore, the frequency spectrum $\mathrm{S}(\omega)$ of a given fluctuation $f(t)$ is given by

$$\mathrm{S}(\omega) = \int_{-\infty}^{\infty}\mathrm{f}(t)\varepsilon^{-j\omega t}dt \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The functions $\mathrm{f}(t)$ and $\mathrm{S}(\omega)$ are Fourier transforms.

Consider a two-dimensional problem where the field in the x-direction is uniform and the angular spectrum is given by $\mathrm{P}(\sin\theta)$. Then the electromagnetic field at a point P may be represented[143] by

$$E_{xy} = \frac{1}{\lambda}\int_{-\infty}^{\infty}\mathrm{P}(\sin\theta)\exp[-j\beta(x\cos\theta + y\sin\theta)]\frac{d(\sin\theta)}{\cos\theta} \quad . \quad (6)$$

$$H_{xy} = \frac{1}{Z\lambda}\int_{-\infty}^{\infty}\mathrm{P}(\sin\theta)\exp -j\beta(x\cos\theta + y\sin\theta)\frac{d(\sin\theta)}{\cos\theta} \quad . \quad (7)$$

where Z is the intrinsic impedance of free space, $\sqrt{(\mu_0/\epsilon_0)}$.

These fields are produced by an aperture distribution

$$Ey = \frac{1}{\lambda}\int_{-\infty}^{\infty}\mathrm{P}(\sin\theta)\exp(-j\beta y\sin\theta)\,d(\sin\theta). \quad . \quad (8)$$

assuming the distance is large enough so that $\cos\theta = 1$. Thus the angular spectrum is expressed in terms of the aperture distribution by the relation

$$\mathrm{P}(\sin\theta) = \int_{-\infty}^{\infty}E_y\exp(j\beta y\sin\theta)dy \quad . \quad . \quad . \quad (9)$$

At a distance from the aperture large compared with its width and the wavelength, the radiation becomes approximately radial and eqn. (9) represents the usual polar diagram in terms of the aperture distribution. The equivalence of eqns. (3) and (9) should be noted.

If eqn. (9) is applied to a rectangular aperture illuminated with a plane-polarized field uniform both in phase and amplitude the angular spectrum becomes the well-known relation

$$\mathrm{P}(\sin\theta) = A_1(\sin\phi)/\phi \quad . \quad . \quad . \quad . \quad (10)$$

where A_1 is a constant and $\phi = (\pi a\sin\theta)/\lambda$. This angular spectrum is plotted as curve (a) in Fig. 2, the values being normalized to give unity at $\theta = 0$.

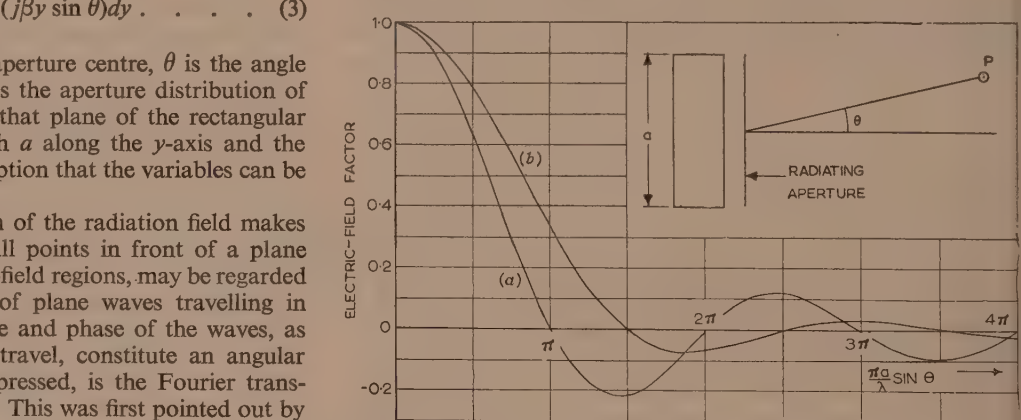


Fig. 2.—Curves for determining radiation fields of an aperture.
The curves give the electric field factor for a radiator with rectangular aperture.
(a) Uniform illumination.
(b) Cosinusoidal illumination.

Similarly a rectangular aperture illuminated with a plane-polarized field uniform in phase but varying as $\cos(\pi y/a)$ in amplitude gives the angular spectrum

$$\mathrm{P}(\sin\theta) = A_2\left[\frac{\sin(\phi + \pi/2)}{\phi + \pi/2} + \frac{\sin(\phi - \pi/2)}{\phi - \pi/2}\right] \quad . \quad (11)$$

where A_2 is a constant. This spectrum is shown as curve (b) in Fig. 2. It will be observed that with this tapered aperture dis-

tribution there is a pronounced reduction in the wide-angle radiation or side lobes but at the expense of a wider main beam.

A more exact formulation of diffraction was made by Kottler:[95] the solutions take into account the vector nature of the electromagnetic field and satisfy Maxwell's equations subject to the prescribed boundary conditions. Sommerfeld[142] gave an exact diffraction theory for reflecting screens which is independent of Huyghens's principle, but its mathematical complexity has prevented its application except in cases of simple geometry such as the infinite half-screen. Exact solutions of finite apertures in plane screens have been given by Andrejeweski[5] for a circular shape and Morse and Rubenstein[113] for an ellipse.

Measurements at microwave frequencies[6, 18, 134] show that the approximate theories based on the Huyghens–Kirchhoff formula provide acceptable results in the cases of apertures large compared with the wavelength and at field points not too close so that polarization effects and edge currents are relatively small. Such approximations have been given for circular apertures,[111, 122] long cylinders,[2] tandem slits,[4] rods,[7] long slits,[77] wedges[72] and other shapes.[33, 80]

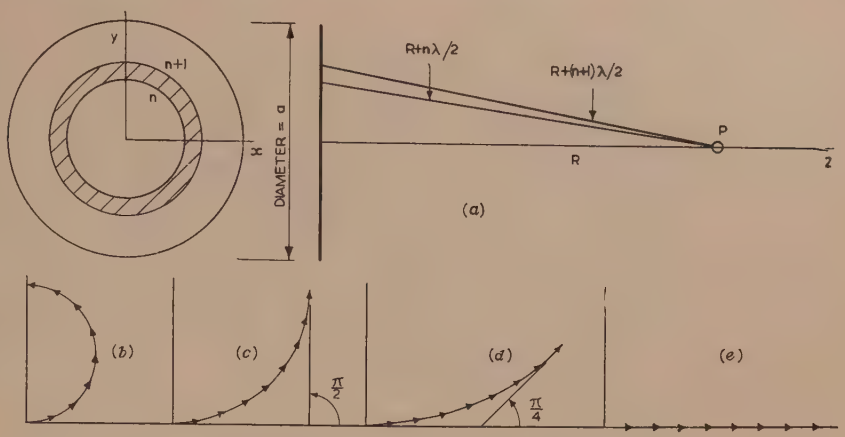


Fig. 3.—Fresnel and Fraunhofer diffraction.

(*a*) Division of a radiating aperture into Fresnel zones for a point P, distance *R*. The transition between the Fresnel and Fraunhofer regions is shown for:
(*b*) $R = a^2/4\lambda$.
(*c*) $R = a^2/2\lambda$.
(*d*) $R = a^2/\lambda$.
(*e*) $R = \infty$.

The scattering of waves by periodic structures such as arrays or gratings is of practical interest. The field of an infinite grating excited by a monochromatic plane wave consists essentially of an infinite set of plane waves: some of these waves are propagating modes and carry energy in specific directions, while the remainder are 'surface waves' which are exponentially damped normal to the plane of the grating. The directions of the spectral modes depends upon the wavelength of the exciting radiation. This diffraction by gratings of elementary sources has received much attention from Trentini[149, 150, 151] and others.[26, 74, 88, 152]

(1.2) Types of Diffraction

The general properties of diffraction have been given, but it is helpful to consider, in an elementary manner, two special cases—Fraunhofer and Fresnel diffraction. The condition for Fraunhofer diffraction is that the radiating source and observation point are at a sufficiently large distance from each other so that the rays originating from the Huyghens secondary wavelets are essentially parallel. The radiation from antennae in free space is usually concerned with this Fraunhofer diffraction.

Fresnel diffraction occurs when the rays from the radiating apertures to the observation point are not parallel, so that the phase difference due to this factor must be taken into consideration. To illustrate this criterion, the field on the axis from a circular aperture illuminated with uniform amplitude and phase will be considered. Thus, from Fig. 3(*a*), it is well known[116] that a series of Fresnel zones may be constructed in the radiating aperture. If the contributions to the field at P due to the zones is *S*, then, if *N* is the number of zones, as *R* increases, *N* decreases so that *S* fluctuates between minimum and maximum values given by

$$S = (S_1 - S_N)/2 \simeq 0 (N \text{ even}) \quad . \quad . \quad . \quad (12)$$

$$S = (S_1 + S_N)/2 \simeq S_1 (N \text{ odd}) \quad . \quad . \quad . \quad (13)$$

These fluctuations cease when, as seen from the point P, the entire aperture consists only of a single Fesnel zone. Simple geometry gives this distance along the axis as

$$R = (a^2 - \lambda^2)/4\lambda \simeq a^2/4\lambda \quad . \quad . \quad . \quad . \quad (14)$$

where *a* is the diameter of the aperture. At this point, as shown in Fig. 3(*b*), the contribution from an elementary area near the edge of the aperture is 180° out of phase with that near the centre. Fig. 3(*c*) gives a vector diagram for a distance $a^2/2\lambda$—the so-called Rayleigh[120] distance—at which the contributions differ in phase by 90°. Figs. 3(*d*) and 3(*e*) show that the diagram at infinity is approximated very closely by that at a^2/λ. Exact calculation shows that the gains at distances of a^2/λ and $2a^2/\lambda$ are respectively 0·94 and 0·99 of that in the Fraunhofer field at infinity.

The application of optical techniques to microwave frequencies often results in working within this Fresnel region, or near field, and analysis of the modified diffraction has been made.[55, 78, 79] It has been shown[14, 15, 114, 139] that the phase across a radiating aperture may be modified so as to introduce focusing. Thus consider an axial field point at a distance such that there is a phase difference between the path lengths from the centre and edges of the aperture. If a quadratic phase distribution with a value at the aperture edges of ϕ and of the correct sign is introduced, a converging system is produced. At the focal point the whole aperture appears to be in phase and the gain at the focus is the same as the Fraunhofer gain of the same aperture distribution with a uniform phase front. It may also readily be shown that the angular radiation pattern in the focal plane is identical with that at infinity.

The intensity distribution at the focus of a lens was given for optical wavelengths by Airy,[3] while more recent studies, at microwave frequencies, have been made by Shinn,[137] Farnell,[60] and Mathews and Cullen.[103] The main feature is a long, narrow, approximately ellipsoidal region of high intensity near the focal point with side lobes of weaker intensity at various distances off the axis. There is a phase change of 180° as the wave goes through the focus: this occurs relatively slowly over a distance of many wavelengths and is associated with an increase in phase velocity in this region. The formation of this focused image may be seen in Fig. 4(*a*), which is due to Bachynski and Bekefi.[10] The axial variation of field in the Fresnel region should be noted.

The focusing of the radiation from an aperture to form an image in the Fresnel region has been studied in unpublished work by R. Meredith as a means of improving the transverse linear definition. Suppose there is a point on the axis which would

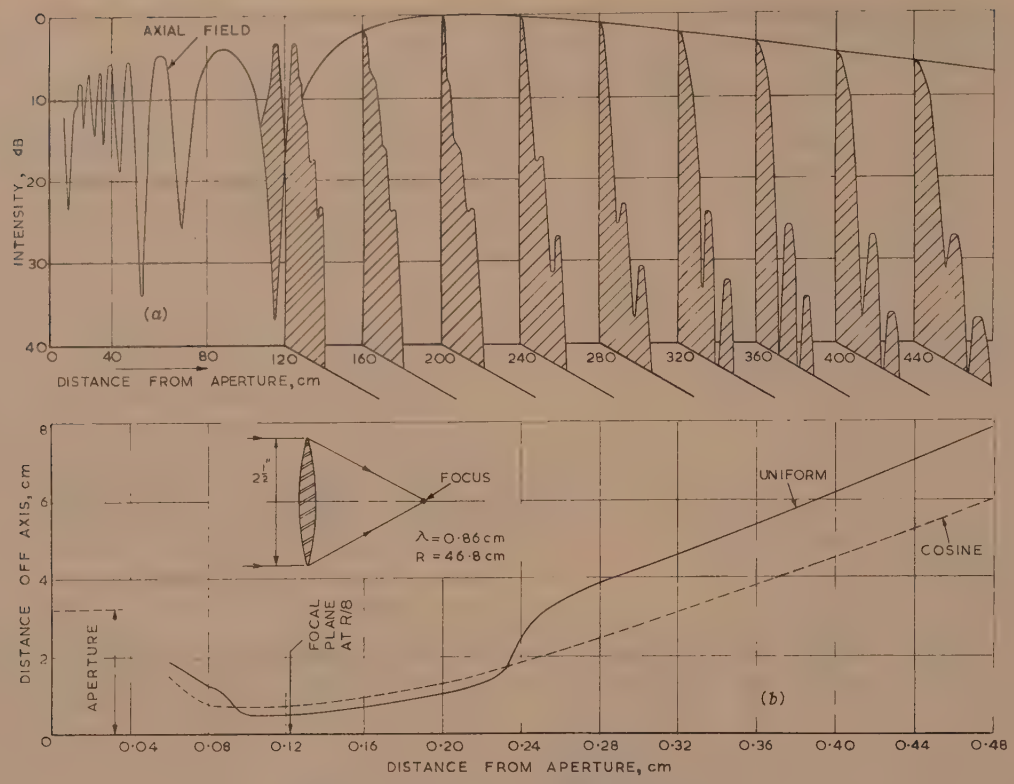


Fig. 4.—Focusing of a radiating aperture.

(*a*) Field distribution in the image space for an ideal radiator, of focal length 360 cm, and Rayleigh distance 120 cm.
(*b*) Radiation contours for an aperture focused at a distance $R/8$, where $R = a^2/\lambda$.

need a quadratic phase ϕ' for focusing. Since the actual distribution is ϕ, there will appear from this point to be a quadratic phase across the aperture of $\phi - \phi'$. The angular radiation pattern at this point is identical with the far-field pattern of the aperture with a quadratic phase of $\phi - \phi'$. This enables a calculation of the transverse linear dimensions of the beam to be made. One example, for a focal length of a quarter of the Rayleigh distance, is given in Fig. 4(*b*). With such a short focal length the depth of field is small and the far-field beam width becomes wide, those of the uniformly illuminated aperture being wider than those of the tampered distribution. The narrowest transverse dimension—or circle of least confusion—occurs some distance from the focal plane and nearer to the aperture. These curves refer to an aperture of only $2\frac{1}{4}$ in square at a frequency of 35 Gc/s, and this effect is likely to be less marked with a wider aperture, i.e. the circle of least confusion would be nearer the focal plane.

(2) ARTIFICIAL DIELECTRICS

(2.1) Obstacle Type

In the application of optical methods and techniques at microwave frequencies, considerable use is made of dielectric media, which may be either of the natural or artificial types. The former, which are of a solid nature, should be chosen to have low loss tangents and to be easy to work: the dielectric constant is always greater than unity. The term 'artificial dielectric' is taken to include any regular array of conducting elements which refracts electromagnetic waves. The general properties of such media have been described by Brown and Jackson[30] and are shown to include permittivity and, in some cases, permeability differing from the free-space values. The refractive index, which may be defined in terms of the reflection coefficient at the interface, may be greater or less than unity. Artificial media may be made inhomogeneous, the refractive index varying from point to point, or anisotropic, the refractive index having different values according to the direction of propagation, or both; in certain conditions they may show[110] the phenomena of Faraday rotation. A number of different artificial media have been developed[32, 38, 63] and can be classified conveniently as obstacle and constrained types.

The delay dielectric of Kock[91] contains a lattice of metallic elements and may be regarded as a large-scale model of a natural dielectric structure. The dielectric constant does not vary appreciably with frequency, provided the dimensions and spacings of the elements are less than, say, one-tenth of a wavelength. If the density, or number of elements per unit volume, is N and the dipole strength, or polarizability, of each elements is α, the effective dielectric constant is given approximately by

$$\epsilon = 1 + (N\alpha/\epsilon_0) \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

A typical medium is a cubical lattice of conducting spheres, shown in Fig. 5(*a*), supported in polystyrene foam. If a is the radius of the spheres, the polarizability is[1]

$$\alpha = 4\pi\epsilon_0 a^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

so that, for static fields, the dielectric constant becomes

$$\epsilon = 1 + 4\pi N a^3 = 1 + (4\pi a^3/s^3) \quad . \quad . \quad . \quad (17)$$

where s is the centre spacing of the spheres. A more exact formula, taking into account the effect of neighbouring spheres, has been given by Lewin.[99] The relative permeability, μ, of the array of spheres differs from unity because of induced eddy

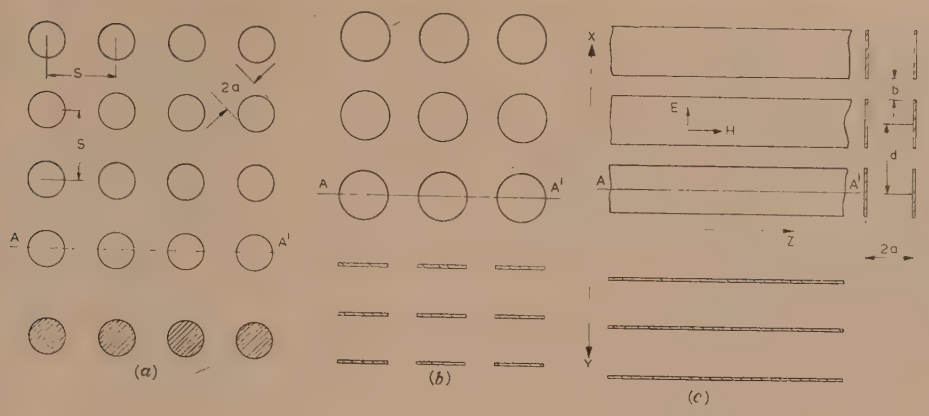


Fig. 5.—Metallic delay dielectrics.

Those shown are lattices of (*a*) spheres, (*b*) discs, and (*c*) strips.
These are all conducting and are supported in foamed polystyrene (not shown). The lower parts of the drawings are sections in planes containing AA′ and normal to the paper.

currents, and Brown[32] has shown that the refractive index n of the medium is given by

$$n^2 = \epsilon\mu = [1 + (4\pi a^3/s^3)][1 - (2\pi a^3/s^3)] \quad . \quad . \quad (18)$$

These relations have been confirmed experimentally at audio frequencies by El-Kharadly and Jackson,[56] and at microwave frequencies by Corkum,[49] using waveguide methods, and El-Kharadly,[57] using parallel-plate methods. For example, a cubical array of spheres with $a = 0{\cdot}3$ cm, $s = 1{\cdot}0$ cm, would have a refractive index of $1{\cdot}12$ and an intrinsic impedance $0{\cdot}75Z$ for wavelengths in the range 7·5–11·0 cm. Such media perform equally well with both horizontally and vertically polarized waves.

The effect of eddy currents can be overcome by using dielectric spheres or metal discs, as shown in Fig. 5(*b*), which have a negligibly small dimension in the direction of propagation. The scattering of discs has been examined by Bouwkamp[24] and Severin,[132, 133] while formulae for the dielectric constant of cubical arrays of discs have been derived by Estrin[58, 59] and Brown and Jackson[31] which agree well with experimental results. The susceptance of disc obstacles was determined experimentally by Cohn[41, 42] with the aid of an electrolytic tank. Lattices of metal flakes have been considered by Swarup,[146] cylindrical and spherical voids by Ward[157] and anisotropic particles by Kaprielian.[87] Scattering of a lattice when the wavelength is less than the particle spacing has been shown by Ramsay and Snook[118] to have close connection with the subject of X-ray crystallography, thus enabling models of particular crystals to be studied at microwave frequencies.

If the medium is to function with only one plane of polarization, Collin[48] and Rytov[127] show that a simple artificial medium may consist of layers of solid dielectric with alternately small and large dielectric constants. Another medium consists of thin metal strips laying at right angles to the electric field and in planes at right angles to the direction of propagation as shown in Fig. 5(*c*). An exact analysis of this medium has been carried out by Cohn,[43, 44, 45] Primich,[117] Bennett,[12] and Brown,[27] the supporting dielectric being assumed to have a dielectric constant the same as that of free space. Since the electric vector is everywhere parallel to the x-direction, perfectly conducting sheets may be inserted along the planes $x = 0$, $x = d$ without disturbing the field. Within the parallel-plate transmission line formed by such sheets, the strips of the metallic delay system form a series of capacitive irises. The dielectric medium may therefore be represented by a transmission line of characteristic impedance $Z_0 d$ loaded with capacitance C, where

$$C = (2\epsilon_0/\pi) \log_e [\mathrm{cosec}\,(\pi b/2d)] \quad . \quad . \quad . \quad . \quad (19)$$

If λ, λ_g are the wavelengths in free space and the delay system respectively,

$$\cos\frac{2\pi a}{\lambda_g} = \cos\frac{2\pi a}{\lambda} - \frac{\omega C Z_0 d}{2}\sin\frac{2\pi a}{\lambda} \quad . \quad . \quad (20)$$

This equation, relating to a periodic structure, defines, as is well known, a series of pass bands, of which only the first is of practical importance. The refractive index can then be shown to be given by

$$n = \lambda/\lambda_g \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

For wavelengths which are large in comparison with the spacing a, the refractive index has the constant value

$$n = \sqrt{\left[1 + \frac{2d}{\pi a}\log_e\left(\mathrm{cosec}\,\frac{\pi b}{2d}\right)\right]} \quad . \quad . \quad . \quad (22)$$

The characteristic impedance is given by

$$Z = Z_0 \tan(\pi a/\lambda)/\tan(\pi a/\lambda g) \quad . \quad . \quad . \quad (23)$$

which reduces to Z_0/n when the ratio a/λ becomes small. A system investigated by Jones and Brown,[85] with $a = 0{\cdot}64$ cm, $b = 0{\cdot}1$ cm, $d = 1{\cdot}0$ cm, gave a limiting refractive index of $1{\cdot}7$ and characteristic impedance of 220 ohms. This analysis is only accurate when the direction of propagation is along or nearly along the z-axis: indeed, it may be shown that in certain directions propagation is not possible, the wave being then exponentially attenuated.

Delay dielectrics are usually constructed from thin insulating sheets of paper or polythene, upon which the conducting strips are mounted, sprayed or printed. These are then spaced apart with Styrofoam, Onazote, Dufaylite or similar expanded materials. Transmission losses for complete media are typically[136] a small fraction of a decibel per metre. For a refractive index of $1{\cdot}5$ the loss tangents are about $0{\cdot}002$ in the wavelength range 3–10 cm but may be even better with very low-loss dielectric supports.

Another type of artificial dielectric contains a lattice structure of perfectly conducting rods as shown in Fig. 6, the rods lying in the direction of the x-axis, the direction of propagation being

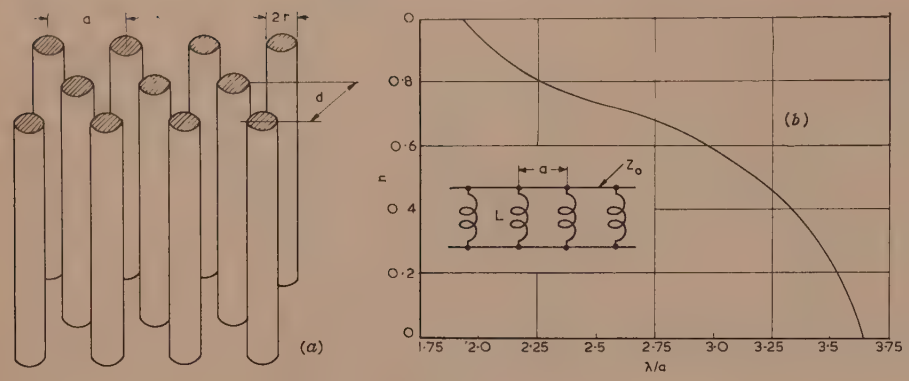


Fig. 6.—Rodded-dielectric medium.

(*a*) Lattice structure.
(*b*) Refractive index.
The inset shows the equivalent circuit of an inductively loaded line.

along the *z*-axis. The structure only functions as a dielectric if the electric field is parallel to the rods and may again be analysed by analogy with a transmission line which in this instance is inductively loaded. The refractive index is given by Brown and Jones[28] as

$$\cos\frac{2\pi an}{\lambda} = \cos\frac{2\pi a}{\lambda} + \frac{\lambda \sin(2\pi a/\lambda)}{2d\log_e(d/2\pi r)} \quad . \quad . \quad (24)$$

where a, d are the lattice spacings and r is the radius of the rods. There is no range of wavelengths for which the refractive index is constant: the rodded medium is dispersive. The refractive index is always less than unity, and the main advantage of this medium is that the rods are strong enough not to require a supporting material. Experimental results by El-Kharadly and Jackson,[56] and El-Kharadly,[57] have shown good agreement with eqn. (24). Typical dimensions for a medium with a refractive of 0·6 at 10·7 cm wavelength, with 0·15 cm-diameter rods are $a = 3{\cdot}55$ cm, $d = 3{\cdot}90$ cm.

(2.2) Constrained Type

It was suggested independently by Rust,[125] Kock,[93] and Stuetzer[144] that an array of conducting plates, as shown in Fig. 7(*a*), would form an artificial dielectric. The plate spacing a is normally chosen so that

$$\lambda/2 < a < \lambda \quad . \quad . \quad . \quad (25)$$

and thus only the TE_{01}-mode can propagate freely between the plates. A rigorous solution of the total field at any point has been given by Carlson and Heins,[35] Berz,[13] and Whitehead.[159] The phase velocity in the medium is given by

$$v_p = c/[1 - (\lambda^2/4a^2)]^{1/2} \quad . \quad . \quad (26)$$

and thus the refractive index is

$$n = [1 - (\lambda^2/4a^2)]^{1/2} \quad . \quad . \quad (27)$$

which is seen to be less than unity. It is found that the reflection and transmission coefficients, referred to the interface, are complex owing to energy stored in the evanescent fields and are modified when the plates have a finite thickness.

Refraction at the free-space/dielectric interface depends on the polarization of the incident wave. If this is in the *E*-plane, as shown in Fig. 7(*a*), then Snell's law of refraction holds. If, however, it is in the *H*-plane, as in Fig. 7(*b*), Snell's law does not necessarily hold, since the waves are constrained to travel only in directions parallel to the plates. The reflection from the metal-plate medium does not exactly obey Fresnel's laws, since the amplitude distribution between a pair of plates is sinusoidal, whereas the incident radiation is uniform.

A symmetrical metal-plate medium, which operates with either direction of plane polarization, is obtained by inserting conducting plates at right angles to the original set as shown in Fig. 7(*c*). Thus waves passing through the medium are constrained in both plates and hence travel parallel to the axis. The refractive index is also, of course, the same for both polarizations.

Eqn. (27) shows that the refractive index varies with wavelength and so these metal-plate media are dispersive. The variation of refractive index n with λ/a is shown in Fig. 7(*d*). Measurements of the reflection coefficients of such media have been made by Primich[115, 116] in the wavelength range 8·0–11·0 cm. Results for the magnitude and phase were obtained while short plates as well as thick and thin ones were examined. Eqn. (27) shows it is possible to achieve any desired variation of refractive index by suitably altering the plate spacing. A region of inhomogeneous refractive index, produced in this way, will therefore consist of a pair of metal plates suitably shaped.

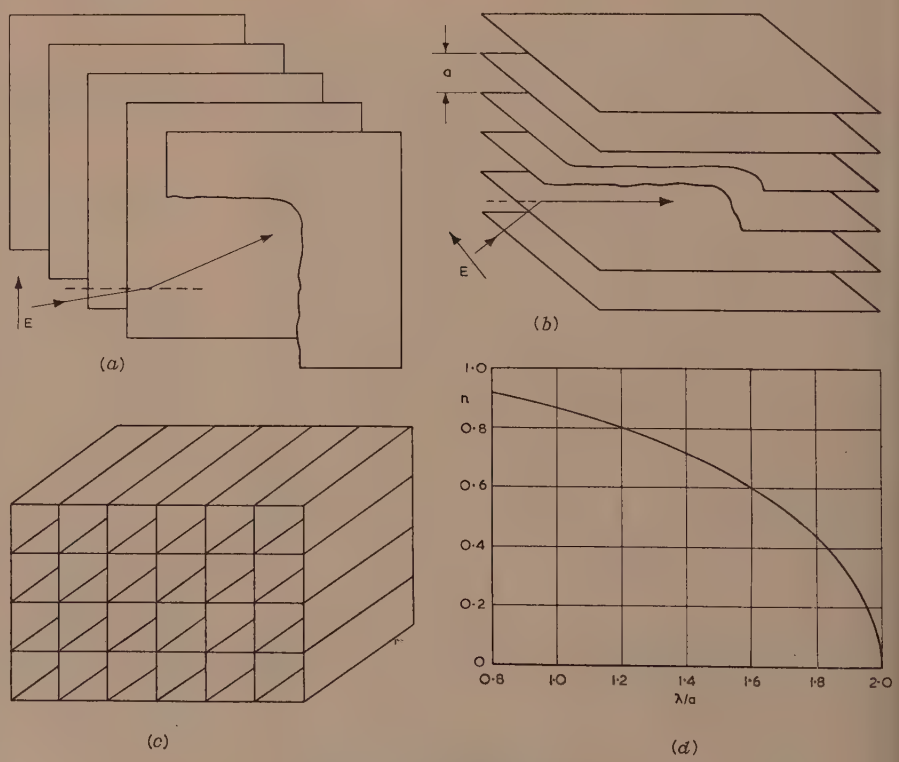


Fig. 7.—Metal-plate media.

The parallel-plate system is shown with the plane of incidence (*a*) parallel and (*b*) perpendicular to the plates.
(*c*) Symmetrical egg-box construction.
(*d*) Dependence of the refractive index on wavelength.

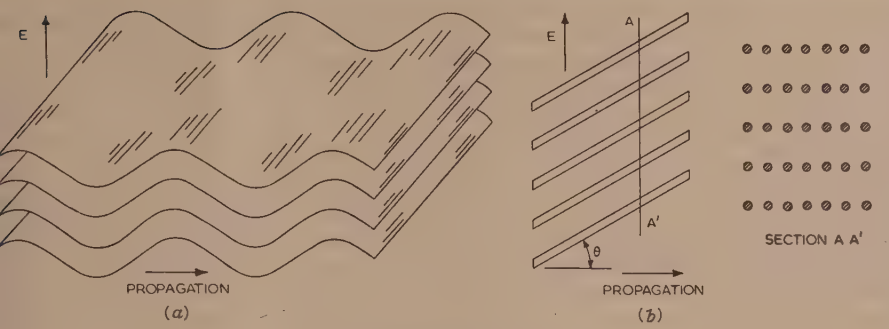


Fig. 8.—Path-length media.
(*a*) Serpentine-plane.
(*b*) Inclined-grid construction.

The path-length medium, introduced by Kock,[92] is another example of a constrained artificial dielectric. The first type, shown in Fig. 8(*a*), consists of a stack of plane parallel conducting sheets, spaced in the direction of the electric field, which is distorted to a serpentine shape by including a number of corrugations. The geometric path length of the energy traversing the medium is seen to be greater than that through the corresponding piece of free space. The medium itself has a refractive index of unity, the dielectric properties depending entirely on phase control by varying the path length of the ray. The medium has an apparent refractive index given by the ratio of the length of the distorted path to that of its axis of symmetry. Provided that the spacing between plates is less than $\frac{1}{2}\lambda$, this refractive index is independent of wavelength and thus the medium has a very good broad-band performance.

The second type of path-length medium is one in which the stack of metal plates is inclined at an angle θ to the direction of the axis. In this case the refractive index is $\sec\theta$. Instead of solid plates, a wire-grid structure can be employed, provided the spacing of the wires is less than $\frac{1}{2}\lambda$. This medium is shown in Fig. 8(*b*), although the wire-grid construction can also be used with the serpentine medium.

(3) CONTROL OF SURFACE REFLECTION

(3.1) Surface Matching

It is well known that an electromagnetic wave incident normally on the interface of a dielectric medium gives rise to a reflected wave whose amplitude is related to the reflection coefficient

$$r = (\epsilon^{1/2} - 1)/(\epsilon^{1/2} + 1) \quad . \quad . \quad . \quad . \quad (28)$$

Such reflected waves result in a direct loss of power from the source of radiation as well as troublesome impedance changes and spurious radiation from apertures. They can be eliminated, as with optical systems, by the application of a dielectric layer, one quarter-wavelength in thickness and of dielectric constant $\sqrt{\epsilon}$, the square root of that of the main body. Such a quarter-wave transformer can be reproduced quite easily with artificial dielectrics simply by reducing the size of the obstacles in the outermost layers. This procedure, suggested by Cohn[45] for the strip medium, makes it possible to obtain a match to free space over a wide range of wavelengths. Reflections from parallel-plate media have been considered by Lengyel[97] and Wells:[158] they may be reduced by stepping the plates a quarter wavelength.

For solid dielectrics the quarter-wave-transformer technique is restricted by a lack of suitable materials. An equivalent effect is obtained, however, by slotting the dielectric surface as shown in Fig. 9(*a*), the slots being directed either parallel or perpendicular to the direction of the electric field in the wave incident on the surface. The slot spacing d should be small enough to ensure that no unwanted propagated waves are excited either in free space or within the dielectric: none will propagate if $d < \lambda/(2\sqrt{\epsilon})$. An analysis of the reflections at the junctions between free space, the slotted dielectric region and the continuous dielectric shows that the behaviour corresponds to abrupt impedance changes at each junction, and that the impedance of each section is directly proportional to the wavelength measured within it. The wavelength λ_s in the slotted section must satisfy the relation

$$\lambda_s = \lambda/\epsilon^{1/4} \quad . \quad . \quad . \quad (29)$$

The depths l of the slots must therefore be given by

$$l = \lambda_s/4 = \lambda/(4\epsilon^{1/4}) \quad . \quad . \quad . \quad . \quad . \quad (30)$$

When the slots are parallel to the electric field the reactive fields excited at each of the junctions may be neglected, and Collin,[47] and Collin and Brown[46] show that the thickness c of the dielectric in the slotted section must satisfy the relation

$$\epsilon^{1/4} \tan \frac{\pi c(\epsilon - \epsilon^{1/2})^{1/2}}{\lambda} = \tanh \frac{\pi(d - c)(\epsilon^{1/2} - 1)^{1/2}}{\lambda} \quad . \quad (31)$$

A set of curves showing c/d against d/λ for various values of ϵ is given in Fig. 9(*b*). A matching layer made to this design

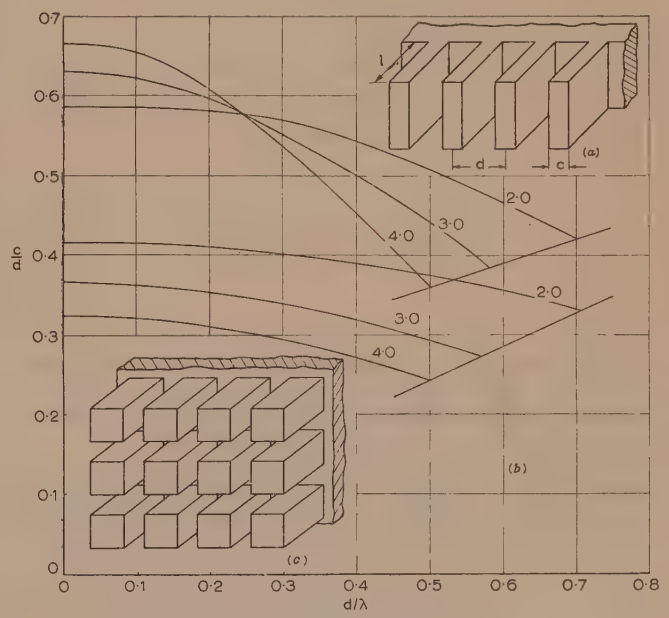


Fig. 9.—Matching layer for a dielectric surface.

(*a*) Slotted medium.
(*b*) Design data for various dielectric constants. The free-space wavelength is λ. The upper and lower sets of curves are for the electric field respectively perpendicular to and parallel with the plane of incidence. The straight lines represent $d/\lambda = 1/\sqrt{\epsilon}$.
(*c*) Double slotted medium.

for polystyrene ($\epsilon = 2{\cdot}56$) reduced the reflected power to less than 0·25% for wavelengths between 9·0 and 11·5 cm. The unmatched surface reflected 5·3% of the incident power.

When the slots are perpendicular to the electric field the

thickness of the dielectric in the slotted section must satisfy the relation

$$\tan \frac{\pi c(\epsilon - \epsilon^{1/2})^{1/2}}{\lambda} = \epsilon^{3/4} \tanh \frac{\pi(d-c)(\epsilon^{1/2}-1)^{1/2}}{\lambda} \quad . \quad (32)$$

A set of design curves is also given in Fig. 9(*b*). With this polarization R. H. Garnham found, in unpublished work, that the reactive fields gave a small residual effect which could be cancelled by reducing the slot depth slightly from the value given in eqn. (32): at 8·6 mm wavelength an amplitude reflection coefficient of 0·01 was obtained with polystyrene.

An obvious way of achieving a matching medium for both orthogonal and circular polarization is to cut two perpendicular sets of slots of equal spacings and widths as shown in Fig. 9(*c*). The computation of such a structure is difficult, so that the dimensions must be found by experiment. It is, however, possible to choose a width of parallel slots which is the average of those required for both polarizations, and a reasonable performance has been obtained.

appears inductive to a plane wave, polarized in the x-direction, incident upon it. If the strips are $2D$ wide with a centre spacing of s, the reactance to a wave polarized in the plane of incidence becomes

$$\frac{X}{Z_d} = \frac{s}{\lambda_d} \cos\theta_d \left[\log_\epsilon \left(\operatorname{cosec} \frac{\pi D}{s} \right) + \mathrm{F}\left(\theta_d, \frac{s}{\lambda_d}, \frac{D}{s}\right) \right] \quad . \quad (35)$$

where $Z_d = 1/Y_d$, θ_d is the angle of incidence and λ_d the wavelength in the dielectric medium. The reactance for a wave polarized perpendicular to the plane of incidence is

$$\frac{X}{Z_d} = \frac{s}{\lambda_d} \cos\theta_d \left[\log_\epsilon \frac{s}{\pi D} + \mathrm{F}'\left(\theta_d, \frac{s}{\lambda_d}\right) \right] \quad . \quad . \quad (36)$$

The correction functions F(—) and F′(—) are small for normal incidence, but the former decreases[101] further while the latter increases[100] for angles off the normal. In Fig. 10(*b*), the single-boundary power-reflection coefficient is plotted for a dielectric with refractive index 1·57 matched at normal incidence by a

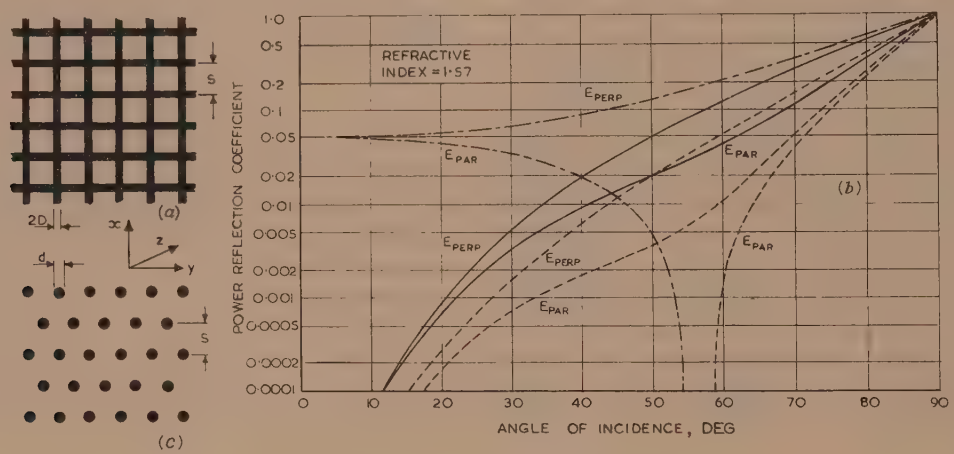


Fig. 10.—Surface matching reactive walls.
(*a*) Inductive strip wall.
(*b*) Reflection coefficients.
—·— No matching.
—— With strips.
---- With discs.
(*c*) Capacitive disc wall.

Morita and Cohn[112] have shown that a satisfactory quarter-wave layer may be made by perturbing the surface of the dielectric. Examples included hemispherical and tetrahedron-shaped dimples, and their theoretical work and experimental result indicated that, roughly speaking, about half the total material must be removed from the surface layer by such shaping.

A rather different approach to this problem of surface matching was adopted by Jones and Cohn,[84] who used reactive walls embedded in the dielectric. The reactive wall took a variety of forms, such as arrays of thin strips[155] or of thin conducting discs. This principle of matching is, of course, very similar to that of stub matching on a transmission line. If Y and Y_d are respectively the characteristic admittances of free space and the dielectric medium, then, if ϕ is the electrical spacing of the reactive wall from the interface,

$$Y_d/Y = \tan^2\phi \quad . \quad . \quad . \quad . \quad . \quad (33)$$

and

$$B/Y_d = \tan\phi \quad . \quad . \quad . \quad . \quad . \quad (34)$$

where B is the susceptance of the reactive wall.

A rectangular array of thin strips, as shown in Fig. 10(*a*), grid of strips spaced $0{\cdot}45\lambda$ between centres. The grid is spaced approximately $\lambda_d/8$ inside the dielectric. The power-reflection coefficient for an unmatched condition is also shown for comparison. An array of discs, as shown in Fig. 10(*c*), has a capacitive reactance and must therefore be embedded about $3\lambda_d/8$ inside the dielectric. The susceptance for a wave polarized perpendicular to the plane of incidence is given by

$$\frac{B}{Y_d} = \frac{16\pi}{\sqrt{3}\,.\,\lambda_d s^2 \cos\theta_d} (M_x - P_z \sin^2\theta_d) \quad . \quad . \quad (37)$$

and for parallel polarization by

$$\frac{B}{Y_d} = \frac{16\pi\cos\theta_d}{\sqrt{3}\,.\,\lambda_d s^2} M_y \quad . \quad . \quad . \quad . \quad . \quad (38)$$

M_x, M_y are the magnetic polarizabilities and P_z the electric polarizability of an aperture the same size as the disc. For a circular aperture of diameter d

$$M_x = M_y = 2P_z = d^3/6 \quad . \quad . \quad . \quad . \quad (39)$$

The reflection coefficient with discs is shown in Fig. 10(*b*).

(3.2) Non-Reflecting Absorbers

If the dielectric medium gives no surface reflection and is lso highly attenuating, a non-reflecting absorber at microwave requencies results. Examples have been described by Schmitt,[129] immons[138] and, more fully, by Severin.[131, 135] In free space ne wave impedance, for a plane electromagnetic wave, is given by

$$Z = \sqrt{(\mu_0/\epsilon_0)} = 377 \text{ ohms} \quad . \quad . \quad . \quad . \quad (40)$$

n all planes of constant phase. Therefore zero reflection is btained, in a wide frequency band, when the input impedance f the absorber system is real and independent of frequency. One simple absorber consisted of a resistance card, whose surface esistivity was equal to Z, placed a quarter of a wavelength from reflecting metal plate. A voltage reflection coefficient better han 0·05 was obtained over a bandwidth of ±5%.

Improved performance was obtained by using two conducting ayers with two equal air-gaps in front of a metal sheet. The oltage reflection-coefficient curve now contained two zeros and ts value was kept below 0·1 within the range of one octave. This construction was extended so that the absorber consisted of lternate layers of a lossless dielectric and thin sheets of poorly onducting material as shown in Fig. 11(*a*). The surface resis-

Fig. 11(*b*), from which it will be seen that the voltage reflection coefficient was below 0·1 over a range of at least three octaves.

Another class of non-reflecting absorber consists of a homogeneous attenuating dielectric layer or sheet. The material should, ideally, possess a characteristic impedance equal to that of free space so that

$$\sqrt{[\mu(1 - j \tan \delta_\mu)/\epsilon(1 - j \tan \delta_\epsilon)]} = 1 \quad . \quad . \quad (41)$$

An efficient absorber should have large and equal values of the dielectric and magnetic loss angles δ_ϵ and δ_μ as well as of dielectric constant ϵ and relative permeability μ, and, moreover, these material constants should be independent of frequency.

A more practicable absorber consists of a layer of thickness d on a metal sheet, the dielectric constant ϵ and loss tangent tan δ being adjusted, as shown by Dällenbach and Kleinsteuber,[54] to give zero reflection at a desired frequency. The voltage reflection coefficient is given by

$$r = \frac{[\tanh 2\pi j(pd/\lambda)] - p}{[\tanh 2\pi j(pd/\lambda)] + p} \quad . \quad . \quad . \quad . \quad (42)$$

where

$$p = \sqrt{[\epsilon(1 - j \tan \delta)]} \quad . \quad . \quad . \quad . \quad (43)$$

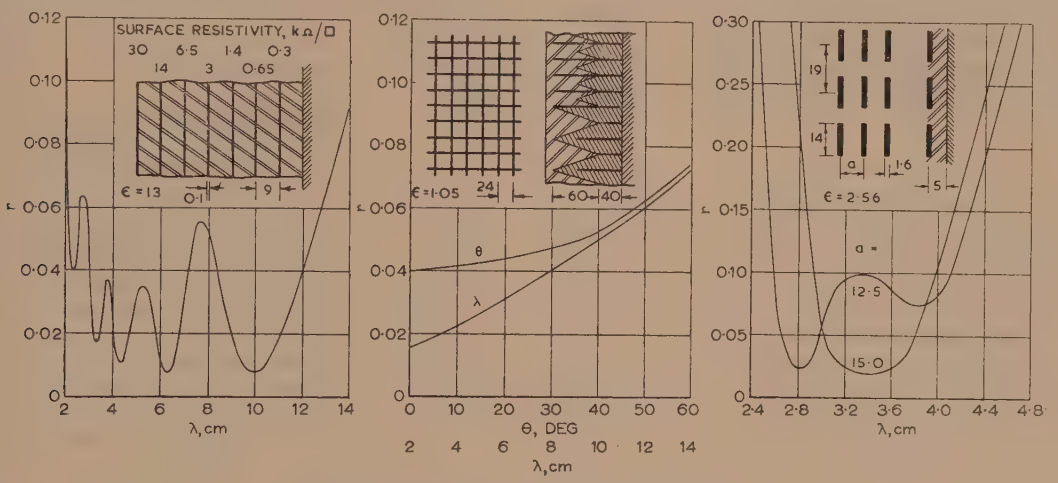


Fig. 11.—Types of non-reflecting absorbers.

ivity decreased towards the metal base by a constant factor from ne sheet to the next, and, as the graph shows, the voltage eflection coefficient calculated by transmission-line theory was elow 0·1 in a range of nearly 3 octaves.

The manufacturing difficulties arising from the exact obserance of specific values of surface resistivity are avoided by sing a uniform absorbing material in such a geometric shape nat the absorption coefficient increases towards the base-plate. n one example, described by Severin,[131, 135] the absorber was uilt up with resistance cards arranged perpendicular to the baseate as illustrated in Fig. 11(*b*). To avoid a dependence upon olarization, two such systems were arranged perpendicular to ach other. Spacings greater than a half-wavelength were used nd the medium corresponded to a highly attenuated waveguide.

The conducting sheets were tapered to match the absorbing arallel-plate medium to the characteristic impedance of free pace, and the teeth were given different lengths to avoid reflecons at sloping angles of incidence owing to the periodic nature f the absorber. The medium was made weatherproof by mbedding the sheets in foamed dielectric. Experimental results or a range of frequencies and angles of incidence are shown in

From the condition for zero reflection

$$p = \frac{\exp[2\pi j(pd/\lambda)] - \exp[-2\pi j(pd/\lambda)]}{\exp[2\pi j(pd/\lambda)] + \exp[-2\pi j(pd/\lambda)]} \quad . \quad . \quad (44)$$

it follows, as a first approximation, for $2\pi d/\lambda < 1$, that the layer thickness must be a quarter of the wavelength in the material so that

$$d = \lambda/4n \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

where the refractive index n is given by

$$n = \sqrt{\left\{\tfrac{1}{2}\epsilon[1 + \sqrt{(1 + \tan^2 \delta)}]\right\}} \quad . \quad . \quad . \quad (46)$$

The surface resistivity has the value

$$1/\sigma d = Z/2 = 188{\cdot}5 \text{ ohms} \quad . \quad . \quad . \quad . \quad (47)$$

where σ is the conductivity of the layer material in mhos/cm. Thus zero reflection is obtained by interference, the amplitudes being adjusted by choice of tan δ and the phase by ϵ and d.

Improved bandwidth can be obtained by multiple layers or by forming the material into a series of pyramids or cones. Suitable

materials have been described by Meyer and others,[107, 109] typical voltage reflection coefficients being less than 0·02 over a frequency range of 2–50 Gc/s.

A magnetically operated absorber may be arranged immediately before a metal sheet, and loss of bandwidth due to path differences is then avoided. For example, magnetic dipoles, in the form of loops, can be arranged to couple to absorbing resonators. In the resonance absorber described by Meyer, Severin and Umlauft,[108] coaxial lines, loaded with iron powder, were used at a frequency of 10 Gc/s. Minimum reflection was obtained when the number of resonators per unit area and their resistance were adjusted to optimum values. Voltage reflection coefficients were typically less than 0·1 for angles of incidence up to 30° and over a frequency range of ±1%.

This bandwidth may be improved by combining two specially dimensioned resonant circuits. A short-circuited line, a quarter of a wavelength long, represents a parallel-resonant circuit, so that, to make a metal plate non-reflecting, suitable elements of series-circuit character have to be mounted at a distance of $\frac{1}{4}\lambda$ in front of the metal plate. Such elements include electric dipoles having the same resonant frequency. The total depth of the absorbing medium may be reduced further by embedding the resonant circuits in a dielectric material. The performance of a typical dipole absorber, due to Severin,[131, 135] is shown in Fig. 11(*c*). With an upper limit of 0·1 for the voltage reflection coefficient, the effective bandwidth extended from $\lambda = 2{\cdot}6$ cm to $\lambda = 4{\cdot}1$ cm with an absorber system having a total depth of 5 mm. Other types of absorption materials have been described by Giger and Tank.[70]

(3.3) Enhancement of Reflectivity

Suitable dielectrics may be used for the converse problem of enhancing the reflectivity of a surface. In comparing the quality of different surfaces, Walker and Hyman[156] considered the ratio of absorbed power to incident power for a TEM-wave. If Z_m is the characteristic impedance of, say, a metal surface, the absorption ratio is given by four times the quotient of the real parts of Z_m and Z, the characteristic impedance of free space. Suppose a sheet of dielectric material of characteristic impedance Z_d, loss factor $\tan\delta$, propagation constant $\gamma = \alpha + j\beta$ and thickness d is spaced $\frac{1}{4}\lambda$ away from the metal surface. The wave impedance at the surface of the dielectric nearest the metal is Z^2/Z_m and thus the wave impedance at the outer surface of the dielectric is

$$Z' = Z_d \frac{(Z^2/Z_m) + Z_d \tanh \gamma d}{Z_d + (Z^2/Z_m)\tanh \gamma d} \quad . \quad . \quad . \quad (48)$$

It can be shown that the real part of Z' is a minimum when d is a quarter-wavelength in the dielectric. Using this value of d and assuming a low-loss dielectric,

$$Z' = (Z_d/Z)^2 Z_m + (\pi/4) Z_d \tan\delta \quad . \quad . \quad . \quad (49)$$

The first term on the right-hand side of this equation takes account of the metal wall and shows that the loss is reduced by the factor $(Z_d/Z)^2$, while the second term depends entirely on the dielectric and accounts for the energy absorbed there.

The condition that the reflection at the outer surface of the dielectric is better than for the metal surface is

$$\tan\delta < (4/\pi Z_d)[1 - (Z_d/Z)^2]\mathscr{R}(Z_m) \quad . \quad . \quad (50)$$

where $\mathscr{R}(Z_m)$ indicates the real part of Z_m. This is also the condition that the reflectivity of an infinite stack of quarter-wave dielectric layers exceeds that of the metal. Evaluation of the impedances in eqn. (50) gives the condition

$$\frac{\sqrt{\epsilon}}{\epsilon - 1}\tan\delta < \frac{\mathscr{R}(Z_m)}{30\pi^2} \quad . \quad . \quad . \quad . \quad (51)$$

so that, for example, dielectric materials for which $[\sqrt{\epsilon}/(\epsilon - 1)]$ $\tan\delta$ is less than 9×10^{-5} should prove better than metals at a frequency of 10 Gc/s.

Evaluation of the impedances in eqn. (49) shows that the wall loss is reduced by a factor R, which, for a single quarter-wave layer of dielectric, is given by

$$R = \frac{1}{\epsilon} + \frac{\tan\delta}{\sqrt{\epsilon}} \frac{30\pi^2}{\mathscr{R}(Z_m)} \quad . \quad . \quad . \quad . \quad (52)$$

or, for an infinity of layers, by

$$R = \tan\delta \frac{\sqrt{\epsilon}}{\epsilon - 1} \frac{30\pi^2}{\mathscr{R}(Z_m)} \quad . \quad . \quad . \quad . \quad (53)$$

The most suitable materials are ceramics having dielectric constants in excess of 10. For example, ceramics consisting mainly of titanium dioxide have a $\tan\delta$ of about 2×10^{-4} and an ϵ of about 90 and give an improvement of about double at 3 Gc/s. The results given in this analysis obtain very closely in the case of waveguides and resonant cavities. An overall reduction in the factor $(\tan\delta)/\sqrt{\epsilon}$ can be obtained by laminating the dielectric normal to the electric field. This is because the energy density is higher in the air spaces than in the dielectric. At microwave frequencies, however, the dimensions of suitable laminations become very small. The bandwidth may be improved by using multiple slightly-detuned layers, but this results in a loss in reflectivity.

(4) INSTRUMENTS AND COMPONENTS

(4.1) Spectrometers

Optical principles can be applied to a number of microwave instruments and components which can be broadly divided into three classes: spectrometers, involving the measurement of angles and directions; interferometers, involving the measurements of length; and beam-dividing devices, involving the measurement of power and attenuation. Since an important limitation is the diffraction which occurs with radiating apertures of practicable dimensions, it cannot, however, always be assumed that the resulting article will perform as efficiently as its optical counterpart. As examples, grids of wires have been used by Valensi[153] as polarization filters, and Kirschbaum and Chen[89] used an anisotropic dielectric to convert plane into circular polarization.

One typical microwave instrument is the analogue of the optical spectrometer, used by Culshaw[53] at a frequency of 24 Gc/s, the arrangement being shown in Fig. 12(*a*). The generator and crystal detector were both connected to horn radiators which were mounted on radial slides rotating about the centre of a graduated circular base one metre in diameter. By means of these slides the distances of the horns from the centre could be varied from 10 cm to 75 cm. The supports which carried the horns were provided with cylindrical graduated heads in which the horns could be rotated about their axes to change the plane of polarization.

Lenses were employed to obtain an aperture distribution uniform in phase but with cosinusoidal and uniform variations of amplitude in the *H*- and *E*-planes respectively. The flare angle was 13°, the aperture 3 in square, and the polystyrene lenses had a focal length of 15 cm and were provided, on both faces, with a non-reflecting layer as well as a 45° bevel on the metal surround. The direct responses, for both perpendicular and parallel polarization, with the horns 38 cm from the centre, are shown in Fig. 12(*b*).

This spectrometer is suitable for measurements of the dielectric constant of various materials in sheet form, the reflectivity being plotted against angle of incidence. Sharp minima were obtained

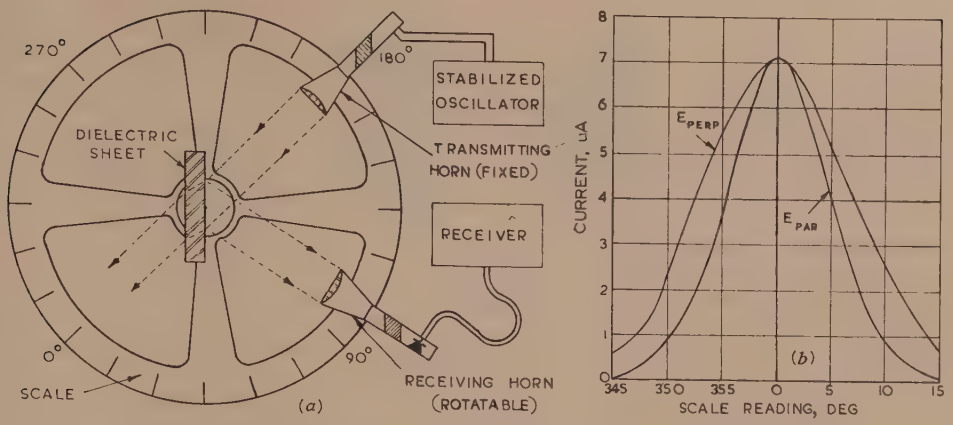


Fig. 12.—Spectrometer for millimetre wavelengths.

(*a*) Layout of the instrument.
(*b*) Direct response with no sample. (Horns at 38 cm; aperture 3 in × 3 in).

ith the horns at a large distance from the centre of the spectro-
ieter, but they became slightly flatter at a distance of about
0 cm. This slight loss of discrimination was associated with the
omewhat larger portion of the angular spectrum which became
ffective in the receiving aperture at spacings within the Fresnel
iffraction region.

Microwave spectrometers have also been described by Goodall
nd Jackson,[71] and Brady, Pearson and Peoples.[25] Roben-
usst[121] used a spectrometer for determining
ie Brewster angle of dielectric materials, and
Cochrane[40] made measurements on prisms of
rtificial dielectric at 3·2 cm wavelength. A
arallel-plate prism of 15° angle was con-
ructed in egg-box fashion, the spacing between
lates being 2 cm to give a refractive index of
·6. The prism aperture was 2 ft square with
ie plates perpendicular to the aperture plane
f the horn on one face. The incident plane
ave was formed by a lens of 2 ft circular
perture, and in the experiments the results
ere found to agree well with the usual theory.

A parallel-plate spectrometer has been de-
cribed by Ruze and Young,[126] and an improved
ersion has been used by Sollom and Brown[141]
or measurements on only small quantities of
rtificial media. The plate spacing was a little
rger than the smaller dimension of the corre-
onding waveguide, making sectoral horns flared in the mag-
etic plane a suitable choice for the feed system. These horns
ad an aperture of 10λ giving a half-power beam width of 7°
nd a maximum side-lobe level 23 dB below that of the main
eam. Such spectrometers have also been used by Seeley[130]
or dielectric measurements.

Spectrometers of a different nature are the echelette grating
pes described by Grillini,[73] Meecham and Peters,[104] and
oates,[39] whose design is shown in Fig. 13. The grating was
rmed by 75 semicircular stainless-steel rods, the flat sides
rming the reflecting surfaces while the feed was a section of a
araboloid. The rods were coupled by a linkage so that as the
rating was rotated about an axis the flat sides always faced the
ed while the effective spacing varied with a cosine law. The
rating was placed about 30 ft from the feed, and measurements
ere made on frequencies within the range 24–50 Gc/s.

(4.2) Interferometers

Interferometers have been used for the measurement of millimetre wavelengths by Martin and Schunemann,[102] Caicoya,[34] and Chatterjee, Bai and Shenoy.[37] An instrument based on the Mach–Zender optical interferometer has been used by Blair[16] for measurements on dielectric sheets at a frequency of 10 Gc/s. One arm of a waveguide bridge included two horns radiating towards each other. Their separation was adjustable and thus enabled the phase shift caused by inserting a sample in the air-gap to be determined. This technique is free from thermal

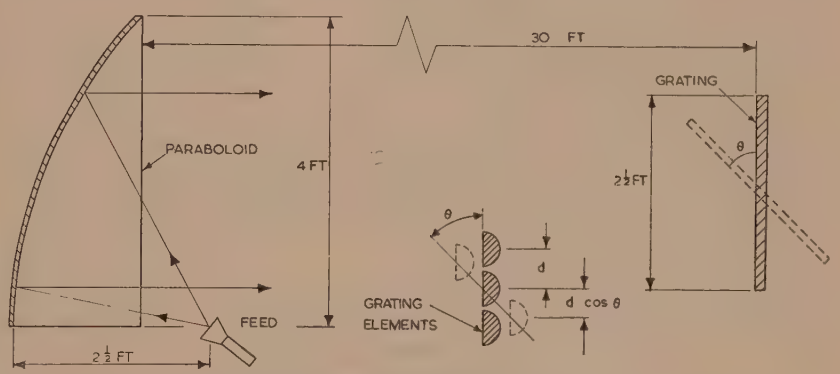


Fig. 13.—Echelette grating spectrometer.

The dimension *d* is 0·4 in, and the normal distance between the elements, *d* cos θ, can be varied by rotating the grating. A linkage mechanism keeps the reflecting faces parallel.

restrictions on the sample and measurements could be made from −72° C to +1 300° C. Using a stabilized generator, phase could be measured to 0·000 2 in.

In the Fabry–Perot interferometer used by Theissing and McGue[148] at 4 mm wavelength, multiple reflections were obtained between two semi-transparent metallic films. A similar instrument used by Artman[9] at 6 mm wavelength comprised one semi-transparent plate based on an artificial dielectric and a metal mirror, the items being spaced so that the radiation was viewed within the first Fresnel zone. In the arrangement of Culshaw,[51] shown in Fig. 14(*a*), multiple quarter-wave sheets of dielectric are used as reflectors. The instrument operates with an essentially parallel beam, and, instead of the circular fringe system of the optical interferometer, a single fringe is observed by altering the distance between the reflectors. The radiation can be made highly monochromatic by frequency stabilization, and the individual fringe shapes can then readily be measured. Three typical fringes at difference spacings of the mirrors are shown in Fig. 14(*b*). The distance between the

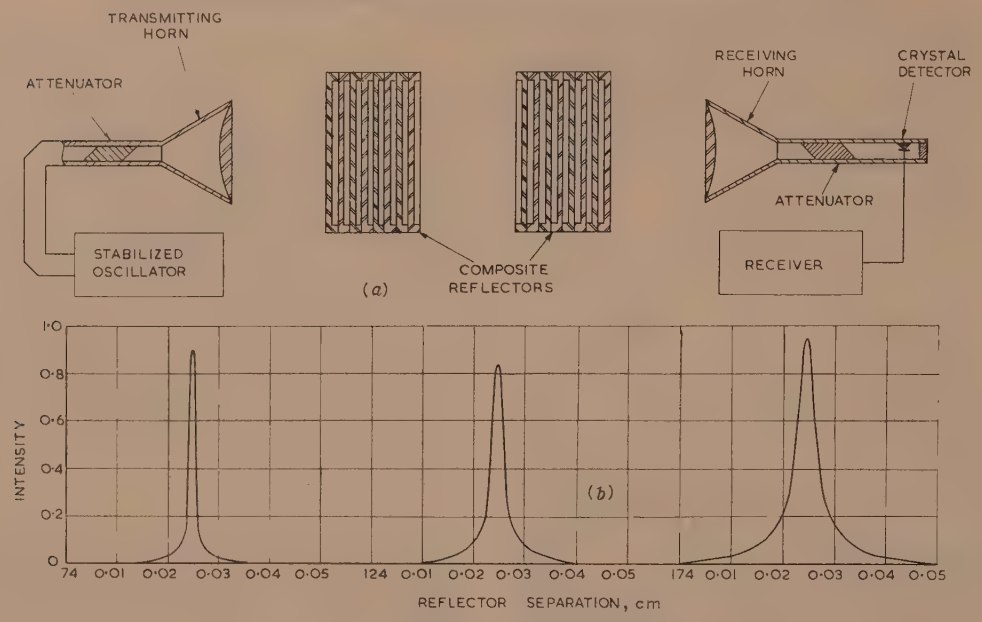


Fig. 14.—Fabry–Perot interferometer.

(*a*) Principle of the interferometer for operation at 35 Gc/s.
(*b*) Responses for three different reflector separations giving Q-factors of 695, 309 and 163.

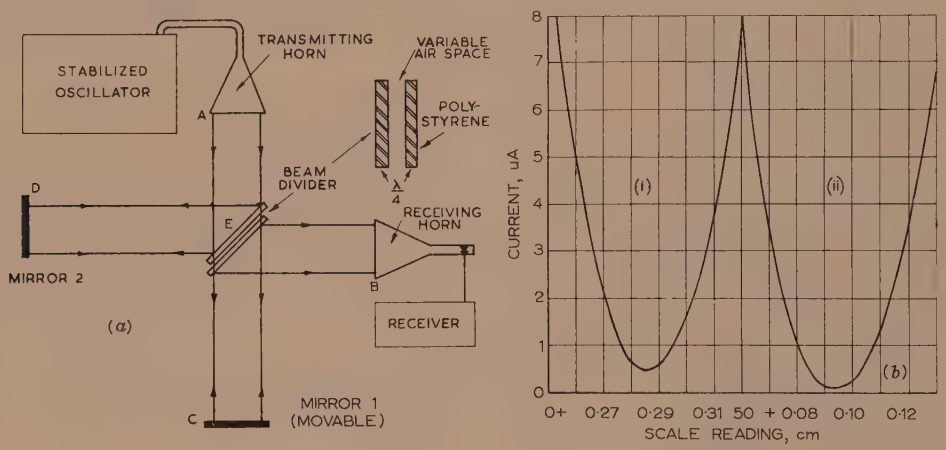


Fig. 15.—Michelson interferometer at millimetre wavelengths.

(*a*) Layout.
(*b*) Response at two positions of the minimum field (AE = BE = 150 cm, DE = 175 cm and CE = (i) 150 cm and (ii) 200 cm).

fringes corresponds accurately to the free-space wavelength, although at close spacings the more pronounced effects of diffraction may cause errors of a few parts in 10^4.

Theory shows that reflectivity is limited by dielectric loss, and although artificial media have been used, materials such as fused quartz are capable of giving reflectivities exceeding that of silver. If a Q-factor is defined for an interferometer as the wavelength divided by the distance between points of half power, Table 1 gives a range of values for different intensity reflectivities. The interferometer described gave values up to 120 000 for Q_n, where n is the order of the interference. The multiple-sheet reflector combines high reflectivity with small transmission loss, thus making the instrument sensitive.

A Boltzmann interferometer was employed by Farrands and Brown[61] for the investigation of the spectrum radiated by spark-type oscillators. This interferometer operated by virtue of the interference of the reflections from two metal plates. The radiation was focused into a plane parallel beam by a paraboloid

Table 1

Q-Factors for Different Intensity Reflectivities

Reflectivity ..	0·50	0·90	0·99	0·999
Q-factor ..	4·5	30	432	2 700

and directed at an angle to the plates. The reflected energy proceeded to another paraboloid and thence to a receiving system. Interference patterns were obtained by displacing one of the plates normal to the plane of its surface.

A true equivalent of the optical Michelson interferometer was used by Culshaw.[50] As shown in Fig. 15(*a*), the energy was radiated from an electromagnetic horn towards a beam splitter. This consisted of two quarter-wavelength plates of low-loss dielectric with an air-gap or spacing adjusted to give a reflection coefficient of 0·5. The transmitted beam was reflected at a fixed

mirror and was further reflected by the beam splitter into a receiving horn and detector. The reflected beam was again reflected at a movable plane mirror and finally transmitted through the beam splitter into the receiving horn.

An interference pattern was set up between the reflected and transmitted beams, and as the movable mirror was continuously traversed along its slide, the detector current became zero at successive half-wavelengths. It was possible to measure a hundred or more such fringes, as shown by the two examples in Fig. 15(*b*), and thus to determine directly the free-space wavelengths.

The arrangement shown was for operation at a frequency of 24 Gc/s, the dimensions of the beam splitter being 9 in wide by 8 in high, and the mirrors were 8 in square and flat to within 0·000 5 in. Wavelength measurements on a source of known frequency gave a velocity of propagation within 1 part in 10^4 of the usually accepted value. This accuracy should be improved by a factor of ten if the effects of atmospheric humidity, pressure and other minor variables are taken into consideration.

Michelson interferometers have also been described by Kahan[86] and Andrews:[8] the true interferometer was modified by Lengyel,[96] who put some of the system in a waveguide. This process was taken a stage further by Froome,[65] who put both the beam splitter or divider and the fixed arm of the interferometer in a waveguide. The arrangement used is shown in Fig. 16, from which it will be seen that the variable arm, i.e. that incorporating the movable reflector, was the only part involving wave transmission in free space. The intensity of the beam in the fixed arm was made to balance that returned from the mirror, so that very sharp interference minima were indicated by the detector. Energy from a frequency-stabilized klystron passed into a hybrid junction which constituted the beam-divider of the interferometer. Half the energy passed through the transmitting horn into the open and was ultimately returned to the horn after reflection from the distant movable mirror. The other half of the energy passed through a matching unit, then a variable attenuator, and was finally reflected by an adjustable short-circuiting plunger. The function of the matching unit was to balance out any unwanted reflections in the free-space arm. In the experiments a sufficiently large distance was used between the horn and the movable reflector for both to be regarded as points sources of radiation. The mirror was displaced through 259 minima, which, for a microwave frequency of 24 Gc/s, corresponded to a distance of 162 cm. After many observations and taking into account many subsidiary sources of error, a value of the velocity of light accurate to 2 parts in 10^6 was obtained.

A symmetrical four-horn interferometer was later used by Froome,[66] the general arrangement being shown in Fig. 17. This symmetry gave improved thermal stability, and by making one arm $\frac{1}{4}\lambda$ longer than the other, the disturbing multiple reflections between each transmitter and the corresponding receiver were arranged to be in opposition. This instrument also works in the Fraunhofer diffraction region, the radiated wavefronts being approximately spheres of radius equal to the separation between transmitter and receiver.

Energy was again divided by a hybrid junction and proceeded to the electromagnetic horns. A phase shifter and variable attenuator were required to adjust and balance the position of the first interference minimum. The plane of polarization of the radiation in the left-hand arm of the instrument was arranged to be at 90° to that in the right-hand arm, thus avoiding cross-coupling between the arms. The instrument was operated by displacement of the carriage through exactly 1 metre, and a variable attenuator, free from phase shift, was used to maintain equal signals in order to obtain sharp minima. An overall accuracy of 1 part in 10^6 was obtained from this interferometer, and a model operating at 72 Gc/s, with less diffraction correction, gave[68] even better performance.

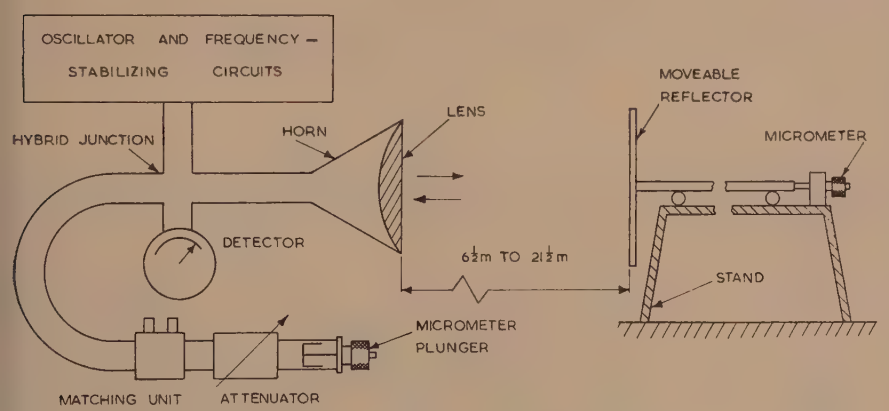


Fig. 16.—Part-waveguide interferometer.
In this arrangement both the beam splitter and the fixed arm of the interferometer are in waveguide.

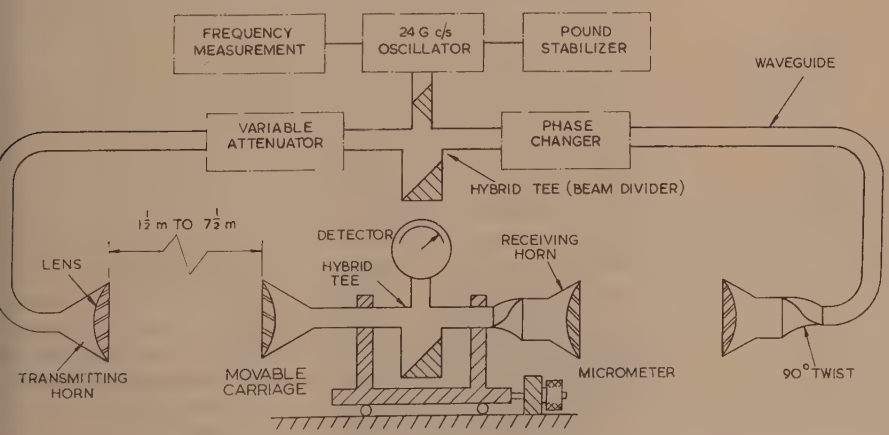


Fig. 17.—Symmetrical four-horn interferometer.

(4.3) Beam-Dividing Devices

Electromagnetic energy incident at, say, 45° upon a wire grid or thin dielectric sheet is divided, one part travelling straight on and the other being reflected at right angles. Beam division may also occur owing to the phenomenon of total reflection exhibited[82] by radiation proceeding from a medium of high to one of low refractive index. Calculation[161] of the electromagnetic fields at the boundary surface shows that they are exponentially attenuated in the space beyond the reflecting surface. If now a second dielectric medium is brought near the first, as shown in the inset of Fig. 18 for the example of two prisms, a certain amount of transmission takes place, as Bose[22] first observed. The ratio of transmitted to reflected power can be varied over a range from zero to a very

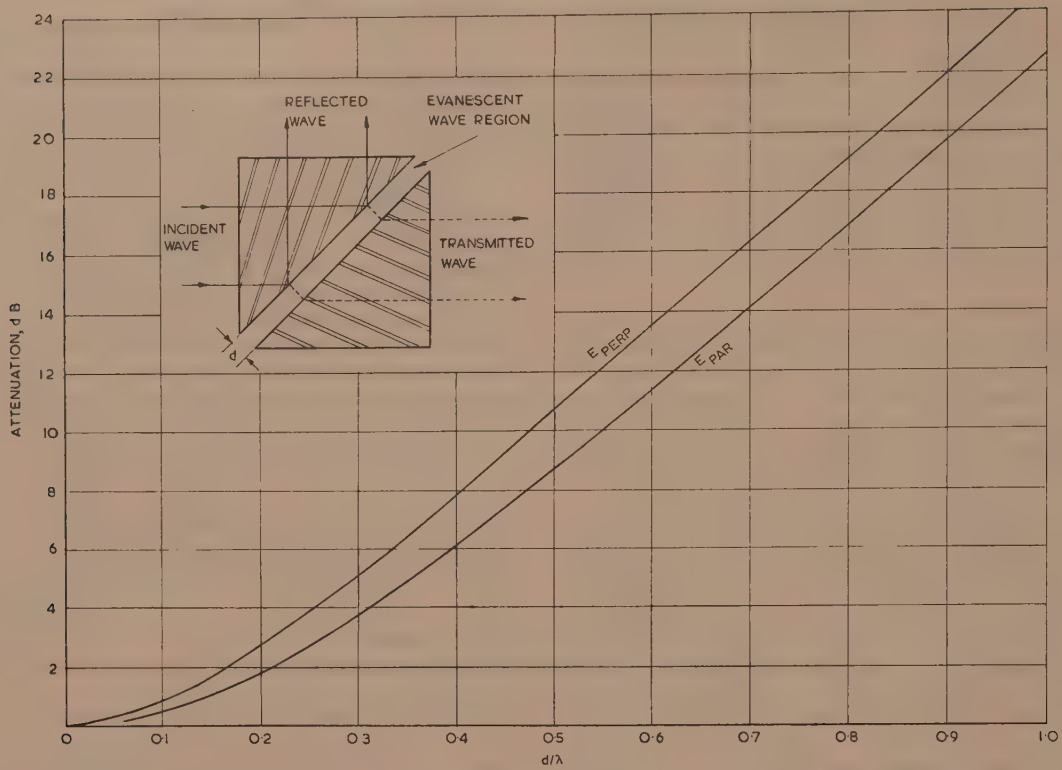


Fig. 18.—Attenuation versus separation of two dielectric prisms.

The curves are for an angle of incidence of 45° and a dielectric constant of 2·545. They are calculated with the electric vector both perpendicular and parallel to the plane of incidence. The slope at larger attenuations is 28·49 dB per wavelength.

large value by adjustment of the distances between the prism faces.

The theory of such an attenuator was derived by Schaefer and Gross.[128] Let θ and ϕ be respectively the angles of incidence and transmission across the air-gap. When θ is greater than the critical angle, Snell's law gives, with due attention to sign,

$$\cos\phi = -j\sqrt{(\epsilon \sin^2\theta - 1)} \quad . \quad . \quad . \quad . \quad (54)$$

Defining an attenuation constant α as

$$\alpha = (2\pi/\lambda)\sqrt{(\epsilon \sin^2\theta - 1)} \quad . \quad . \quad . \quad . \quad (55)$$

the fractional reflected and transmitted amplitudes, E_r/E_i and E_t/E_i respectively, may be calculated for an air-gap d. They are, for the electric field perpendicular to the plane of incidence,

$$\frac{E_r}{E_i} = \frac{(\epsilon - 1)\sinh \alpha d}{(\epsilon + 1 - 2\epsilon \sin^2\theta)\sinh \alpha d - 2j\cos\theta[\epsilon(\epsilon \sin^2\theta - 1)]^{1/2}\cosh \alpha d} \quad . \quad . \quad . \quad . \quad (56)$$

$$\frac{E_t}{E_i} = \frac{-2j\cos\theta[\epsilon(\epsilon \sin^2\theta - 1)]^{1/2}}{(\epsilon + 1 - 2\epsilon \sin^2\theta)\sinh \alpha d - 2j\cos\theta[\epsilon(\epsilon \sin^2\theta - 1)]^{1/2}\cosh \alpha d} \quad . \quad . \quad . \quad . \quad (57)$$

and for the electric field parallel to the plane of incidence

$$\frac{E_r}{E_i} = \frac{(\epsilon - 1)(\epsilon \sin^2\theta - \cos^2\theta)\sinh \alpha d}{[(\epsilon + 1) - (\epsilon^2 + 1)\sin^2\theta]\sinh \alpha d - 2j\cos\theta[\epsilon(\epsilon \sin^2\theta - 1)]^{1/2}\cosh \alpha d} \quad . \quad . \quad . \quad . \quad (58)$$

$$\frac{E_t}{E_i} = \frac{2j\cos\theta[\epsilon(\epsilon \sin^2\theta - 1)]^{1/2}}{[(\epsilon + 1) - (\epsilon^2 + 1)\sin^2\theta]\sinh \alpha d - 2j\cos\theta[\epsilon(\epsilon \sin^2\theta - 1)]^{1/2}\cosh \alpha d} \quad . \quad . \quad . \quad . \quad (59)$$

Calculated values of reflected and transmitted energy are given in Fig. 18 for the case of $\theta = 45°$ and $\epsilon = 2{\cdot}545$. The effect of dielectric loss can be estimated from eqns. (56)–(59), but for loss tangents of 0·001 it is usually quite negligible. In unpublished work, R. H. Garnham has shown that these values of attenuation given by simple theory must be corrected for the effects of diffraction consequent upon the finite aperture sizes. If the polar diagram of the radiation from the transmitting aperture is plotted as an angular spectrum together with values of attenuation at different angles, it becomes evident that waves having smaller angles of incidence are attenuated less than those of larger angles. These effects introduce second- and third-order terms, which, at an attenuation of 50 dB, require corrections of −0·9 dB and −0·1 dB respectively.

Eqns. (56)–(59) show that, for both polarizations of the electric field, the transmitted and reflected waves are in phase quadrature. A. W. Lines showed in unpublished work that the double prism assembly, used in Fig. 19(*a*), is the optical equivalent of the conventional microwave directional coupler and can form the basis of several practical components. Thus the device may be used in conjunction with reflecting surfaces to form an absorption wavemeter. In Fig. 19(*b*) some of the incident power is reflected by the directional coupler into the interferometer consisting of one composite reflector and one metal mirror. As the latter is moved there occur the usual half-wave positions at which the energy reaching the detector drops sharply. The various items

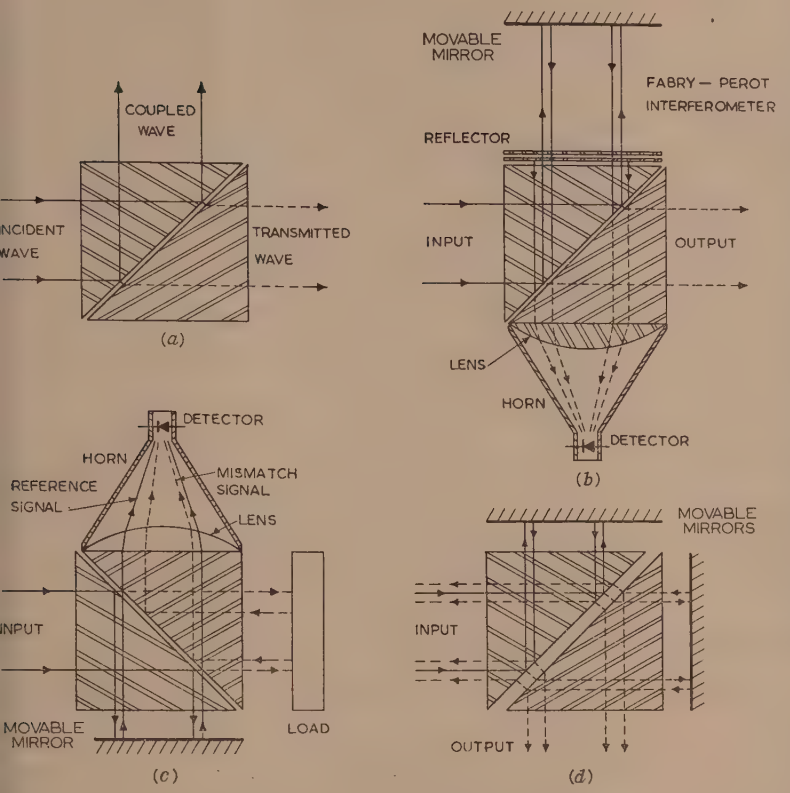


Fig. 19.—Quasi-optical components using prisms.

(*a*) Directional coupler.
(*b*) Absorption wavemeter.
(*c*) Standing-wave meter.
(*d*) *EH* matching unit.

re within the Rayleigh 'cylinder', the horn feed giving, in the ransverse plane, a nearly half-sinusoidal field distribution within ne dielectric with an exponential fall of amplitude outside it. 'ypical dimensions of the prisms would be 6 in each side for a requency of 35 Gc/s.

In the standing-wave meter of Fig. 19(*c*), a small fraction of ne incident power is reflected at the prism gap towards the novable mirror, whence it is reflected towards the detector. Iere it is compared directly with the same fraction of the power eflected from the load. Movement of the mirror gives succes-ive maxima and minima from which the reflection coefficient nay be calculated in the usual way.

If the prism spacing is adjusted to give 3 dB attenuation, a evice having the properties of the usual hybrid junction is btained. In Fig. 19(*d*), movable mirrors placed in the side-rms lead to the analogue of the conventional *EH* hybrid match-ng unit. Auxiliary components such as loads of either very ow or very high reflectivity can be designed in accordance with ne principles discussed earlier, and, if preferred, the assembly ould possibly include artificial dielectrics.

A transducer from plane to circular polarization was made by ulshaw and Jones,[52] a metal reflecting plate being placed at the orrect distance from the hypotenuse of a single prism. The ange of phase consequent upon total internal reflection led . Culshaw, in unpublished work, to construct the equivalent of ne Fresnel rhomb. The phase change δ which appears on total eflection between components in and perpendicular to the plane f incidence is given[143] by

$$\tan(\delta/2) = \sqrt{[\sin^2\theta_0 - (1/\epsilon)^2]}\cos\theta_0/\sin^2\theta_0 \quad . \quad (60)$$

where θ_0 is the angle of incidence on the totally reflecting surface. If this phase difference can be made equal to $\frac{1}{2}\pi$, circular polarization is obtained from an incident plane wave. For polystyrene, $\epsilon = 2{\cdot}5$, two reflections were found necessary, the required angle θ_0 being 58° 04′. Since the only term dependent on frequency in eqn. (60) is the dielectric constant, the Fresnel rhomb should prove a broad-band device.

In view of the low power density, optical-type components are inherently capable of withstanding high peak powers. This fact is especially advantageous at millimetre wavelengths and enabled the author to construct a mismatch unit for 35 Gc/s. The device consisted of two horn radiators tapering from 0·28 in × 0·14 in waveguide. The mismatch reflection was caused by an intervening transverse polystyrene sheet. One of the faces was slotted to reduce frequency sensitivity and obtain the correct reflection coefficient, the other being left flat. The input guide and horn were pressurized and the output horn contained a water load for measuring the power of a generator. Longitudinal movement of the dielectric sheet gave a mismatch of varying phase but constant amplitude.

(5) ACKNOWLEDGMENTS

The author is indebted to Dr. A. W. Lines, Dr. W. Culshaw, Dr. J. Brown, Mr. R. Meredith and Mr. R. H. Garnham for much useful discussion on the subject of optical techniques at microwave frequencies and for information prior to publication, and to Mr. R. H. Garnham for helpful comments on the manuscript.

(6) BIBLIOGRAPHY

(1) ADEN, A. L.: 'Electromagnetic Scattering from Spheres with Sizes Comparable to the Wavelength', *Journal of Applied Physics*, 1951, **22**, p. 601.
(2) ADEY, A. W.: 'Diffraction of Microwave by Long Metal Cylinders', *Canadian Journal of Physics*, 1955, **33**, p. 407.
(3) AIRY, G. B.: 'On the Diffraction of an Object Glass with a Circular Aperture', *Transactions of the Cambridge Philosophical Society*, 1834, **5**, p. 283.
(4) ALLDREDGE, L. R.: 'Diffraction of Microwaves by Tandem Slits', *Transactions of the Institute of Radio Engineers*, 1956, **AP-4**, No. 4, p. 640.
(5) ANDREJEWSKI, W.: 'Rigorous Theory of the Diffraction of Plane Electro-magnetic Waves at a Perfectly Conducting Circular Disc and a Circular Aperture on a Perfectly Conducting Plane Screen', *Die Naturwissenschaften*, 1951, **38**, p. 406.
(6) ANDREWS, C. L.: 'Diffraction Pattern in a Circular Aperture Measured in the Microwave Region', *Journal of Applied Physics*, 1950, **21**, p. 761.
(7) ANDREWS, C. L.: 'Diffraction Pattern of Microwaves near Rods', *ibid.*, 1951, **22**, p. 465.
(8) ANDREWS, C. L.: *American Journal of Physics*, 1946, **14**, p. 379.
(9) ARTMAN, J. O.: 'A Microwave Fabry–Perot Interferometer', *Review of Scientific Instruments*, 1953, **24**, p. 873.
(10) BACHYNSKI, M. P., and BEKEFI, G.: 'Aberrations in Circularly Symmetric Microwave Lenses', *Transactions of the Institute of Radio Engineers*, 1956, **AP-14**, No. 3, p. 412.
(11) BAKER, B. B., and COPSON, E. T.: 'The Mathematical Theory of Huyghens Principle' (Clarendon Press, Oxford, 1950).
(12) BENNETT, H. S.: 'The Electromagnetic Transmission Characteristics of the Two-Dimensional Lattice Medium', *Journal of Applied Physics*, 1953, **24**, p. 785.
(13) BERZ, F.: 'Reflection and Refraction of Microwaves at a Set of Parallel Metallic Plates', *Proceedings I.E.E.*, Paper No. 1038 R, January, 1951 (**98**, Part III, p. 47).
(14) BICKMORE, R. W.: 'On Focusing Electromagnetic Radiation', *Canadian Journal of Physics*, 1957, **35**, p. 1292.
(15) BICKMORE, R. W.: 'Fraunhofer Pattern Measurement in the Fresnel Region', *ibid.*, 1957, **35**, p. 1299.
(16) BLAIR, G. R.: 'An Ultra-Precise Microwave Interferometer', *Convention Record of the Institute of Radio Engineers*, 1958, Part I, p. 48.

(17) BOCHENEK, K., and PLEBEANSKI, J.: 'Methods of Microwave Optics', *Archiwum Elektrotechniki*, 1956, **5**, p. 293.
(18) BOIVIN, A., DION, A., and KOENIG, H. P.: 'Experimental Study of the Diffraction of Microwaves by Rotationally Symmetrical Apertures', *Canadian Journal of Physics*, 1956, **34**, p. 166.
(19) BOLTZMANN, L.: 'Investigation into Hertzian Phenomena', *Annalen der Physik*, 1890, **40** (Series 3), p. 399.
(20) BOOKER, H. G., and CLEMMOW, P. C.: 'The Concept of an Angular Spectrum of Plane Waves, and its Relation to that of Polar Diagram and Aperture Distribution', *Proceedings I.E.E.*, Paper No. 922 R, January, 1950 (**97**, Part III, p. 11).
(21) BOSE, J. C.: 'On the Determination of the Wavelength of Electric Radiation by Diffraction Grating', *Philosophical Magazine*, 1896, **60**, p. 167.
(22) BOSE, J. C.: 'Collected Physical Papers' (Longmans Green, London, 1927).
(23) BOUWERS, A.: 'Achievements in Optics' (Elsevier, New York–Amsterdam, 1946).
(24) BOUWKAMP, C. J.: 'On the Diffraction of Electromagnetic Waves by Small Circular Discs and Holes', *Philips Research Reports*, 1950, **5**, p. 401.
(25) BRADY, J. J., PEARSON, M. D., and PEOPLES, S. R.: 'Reflection of Microwaves from Metal-Plate Media', *Journal of Applied Physics*, 1952, **23**, p. 964.
(26) BROUSSARD, G.: 'Study of the Diffraction of Electromagnetic Waves by an Array of Perforated Plates', *Annales de Radio-électricité*, 1955, **10**, p. 42.
(27) BROWN, J.: 'The Design of Metallic Delay Dielectrics', *Proceedings I.E.E.*, Paper No. 915 R, January, 1950 (**97**, Part III, p. 45).
(28) BROWN, J., and JONES, S. S. D.: 'Microwave Lenses', *Electronic Engineering*, 1950, **22**, pp. 127, 183, 227, 264, 368 and 429.
(29) BROWN, J.: 'Artificial Dielectrics having Refractive Indices less than Unity', *Proceedings I.E.E.*, Monograph No. 62 R, May, 1953 (**100**, Part IV, p. 51).
(30) BROWN, J., and JACKSON, W.: 'The Properties of Artificial Dielectrics at Centimetre Wavelengths', *ibid.*, Paper No. 1699 R, January, 1955 (**102** B, p. 11).
(31) BROWN, J., and JACKSON, W.: 'The Relative Permittivity of Tetragonal Arrays of Perfectly Conducting Thin Discs', *ibid.*, Paper No. 1702 R, January, 1955 (**102** B, p. 37).
(32) BROWN, J.: 'Microwave Lenses' (Methuen, London, 1953).
(33) BUCHSBAUM, S. J., MILNE, A. R., HOGG, D. C., BEKEFI, G., and WOONTON, G. A.: 'Microwave Diffraction by Apertures of Various Shapes', *Journal of Applied Physics*, 1955, **26**, p. 706.
(34) CAICOYA, J. I.: 'The Generation of Multiple TE_{mo} and TM_{mo} Modes between Parallel Plates and in Rectangular Waveguides by Interference and its Application to the Measurement of Wavelength', *Revista de Telecomunicación* (*Madrid*), 1956, **11**, No. 43, p. 2.
(35) CARLSON, J. F., and HEINS, A. E.: 'The Reflection of an E.M. Wave by an Infinite Set of Plates', *Quarterly of Applied Mathematics*, 1947, **4**, p. 313; 1947, **5**, p. 82; and 1950, **8**, p. 281.
(36) CARRARA, N.: 'Refraction of Evanescent Waves', *Alta Frequenza*, 1950, **19**, p. 164.
(37) CHATTERJEE, S. K., and BAI, C. R., and SHENOY, P. R.: 'Microwave Interferometer', *Journal of the Indian Institute of Science*, 1954, **36** B, p. 172.
(38) CHATTERJEE, S. K., and RAO, B. V.: 'Investigations on Artificial Dielectrics at Microwave Frequencies: Part 1', *ibid.*, 1955, **37** B, p. 304.
(39) COATES, R. J.: 'A Grating Spectrometer for Millimetre Waves', *Review of Scientific Instruments*, 1948, **19**, p. 586.
(40) COCHRANE, C. A.: 'An Experimental Verification of the Theory of Parallel Plate Media', *Proceedings I.E.E.*, Paper No. 896 R, March, 1950 (**97**, Part III, p. 72).
(41) COHN, S. B.: 'Electrolytic-Tank Measurements for Microwave Metallic Delay-Lens Media', *Journal of Applied Physics*, 1950, **21**, p. 674.
(42) COHN, S. B.: 'The Electric and Magnetic Constants of Metallic Delay Containing Obstacles of Arbitrary Shape and Thickness', *ibid.*, 1951, **22**, p. 628.
(43) COHN, S. B.: 'Experimental Verification of the Metal Strip Delay Lens Theory', *ibid.*, 1953, **24**, p. 839.
(44) COHN, S. B.: 'Microwave Measurement on Metallic Delay Media', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 1177.
(45) COHN, S. B.: 'Analysis of the Metal Strip Delay Structure for Microwave Lenses', *Journal of Applied Physics*, 1949, **20**, p. 257.
(46) COLLIN, R. E., and BROWN, J.: 'The Design of Quarter-Wave Matching Layers for Dielectric Surfaces', *Proceedings I.E.E.*, Monograph No. 149 R, September, 1955 (**103** C, p. 153).
(47) COLLIN, R. E.: 'Reflection and Transmission at a Slotted Dielectric Interface', *Canadian Journal of Physics*, 1956, **34**, p. 398.
(48) COLLIN, R. E.: 'A Simple Artificial Anisotropic Dielectric Medium', *Transactions of the Institute of Radio Engineers*, 1958, **MTT-6**, No. 2, p. 206.
(49) CORKUM, R. W.: 'Isotropic Artificial Dielectric', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 574.
(50) CULSHAW, W.: 'The Michelson Interferometer at Millimetre Wavelengths', *Proceedings of the Physical Society*, 1950, **63**, B, p. 393.
(51) CULSHAW, W.: 'The Fabry–Perot Interferometer at Millimetre Wavelengths', *ibid.*, 1953, **66** B, p. 597.
(52) CULSHAW, W., and JONES, D. S.: 'Effect of the Metal Plate on Total Reflection', *ibid.*, 1953, **66** B, p. 859.
(53) CULSHAW, W.: 'A Spectrometer for Millimetre Wavelengths', *Proceedings I.E.E.*, Paper No. 1445 M, February, 1953 (**100**, Part IIA, p. 5).
(54) DÄLLENBACH, W., and KLEINSTEUBER, W.: 'Reflection and Absorption of Decimetre-Waves by Plane Dielectric Layers', *Hochfrequenztechnik und Electroakustik*, 1938, **51**, p. 152.
(55) EHRLICH, M. J., SILVER, S., and HELD, G.: 'Studies of the Diffraction of Electromagnetic Waves by Circular Apertures and Complementary Obstacles: the Near-Zone Field', *Journal of Applied Physics*, 1955, **26**, p. 336.
(56) EL-KHARADLY, M. M. Z., and JACKSON, W.: 'The Properties of Artificial Dielectrics Comprising Arrays of Conducting Elements', *Proceedings I.E.E.*, Paper No. 1472 R, May, 1953 (**100**, Part III, p. 199).
(57) EL-KHARADLY, M. M. Z.: 'Some Experiments on Artificial Dielectrics at Centimetre Wavelengths', *ibid.*, Paper No. 1700 R, January, 1955 (**102** B, p. 17).
(58) ESTRIN, G.: 'The Effects of Anistropy in a Three-Dimensional Array of Conducting Disks', *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 821.
(59) ESTRIN, G.: 'The Effective Permeability of an Array of Thin Conducting Disks', *Journal of Applied Physics*, 1950, **21**, p. 667.
(60) FARNELL, G. W.: 'Calculated Intensity and Phase Distribution in the Image Space of a Microwave Lens', *Canadian Journal of Physics*, 1957, **35**, p. 777.
(61) FARRANDS, J., and BROWN, J.: "Boltzmann Interferometer', *Wireless Engineer*, 1954, **31**, p. 81.
(62) FEL'D, YA. N.: 'Incidence of Electromagnetic Waves on Double Infinite Grids', *Comptes Rendus de l'Acedémie des Sciences de l'U.R.S.S.*, 1956, **107**, p. 71.
(63) FORTET, R.: 'Propagation in an Artificial Dielectric', *Annales des Télécommunication*, 1953, **8**, p. 361.
(64) FRANZ, W.: 'Theory of the Trentini Absorption Grid', *Zeitschrift für angewandte Physik*, 1954, **6**, p. 449.
(65) FROOME, K. D.: 'Determination of the Velocity of Short Electromagnetic Waves by Interferometry', *Proceedings of the Royal Society*, A, 1952, **213**, p. 123.
(66) FROOME, K. D.: 'Investigation of a New Form of Microwave Interferometer for Determining the Velocity of Electromagnetic Waves', *ibid.*, 1954, **223**, p. 195.
(67) FROOME, K. D.: "Microwave Determination of the Velocity of Light', *Journal of the British Institution of Radio Engineers*, 1956, **16**, p. 497.
(68) FROOME, K. D.: 'Precision Determination of the Velocity of Electromagnetic Waves', *Nature*, 1958, **181**, p. 258.
(69) GIBELLATO, S.: 'Total Reflection of an Electromagnetic Wave with Metallic Reflection of the Evanescent Wave', *Il Nuovo Cimento*, 1949, **6**, p. 344.
(70) GIGER, A., and TANK, F.: 'Investigation of Radar Absorption Materials', *Schweizer Archiv für angewandte Wissenschaft und Technik*, 1956, **22**, p. 414.
(71) GOODALL, E. G. A., and JACKSON, J. A. C.: 'A Focused Spectrometer for Microwave Measurements', *Marconi Review*, 1957, **20**, p. 51.
(72) GRANNEMANN, W. W., and WATSON, R. B.: 'Diffraction of Electromagnetic Waves by a Metallic Wedge of Acute Dihedral Angle', *Journal of Applied Physics*, 1955, **26**, p. 392.
(73) GRILLINI, L.: 'An Experiment on a Plane Microwave Diffraction Grating', *Il Nuovo Cimento*, 1951, **8**, p. 952.
(74) GROVES, W. E.: 'Transmission of Electromagnetic Waves Through Pairs of Parallel-Wire Grids', *Journal of Applied Physics*, 1953, **24**, p. 845.
(75) HADDENHORST, H. G.: 'Transmission of Electromagnetic Waves through Inhomogeneous Layers', *Zeitschrift für angewandte Physik*, 1955, **7**, p. 487.
(76) HADDENHORST, H. G.: 'Transmission of Electromagnetic Waves through Inhomogeneous Layers: Part 2—Absorption of Electromagnetic Waves', *ibid.*, 1956, **8**, p. 264.
(77) HIRSHFIELD, J. L., and ZIEMAN, C. M.: 'Measurement of Microwave Diffraction from a Long Slit in a Thin Conducting Plane', *Journal of Applied Physics*, 1955, **26**, p. 135.
(78) HOGG, D. C.: 'The Electromagnetic Field in the Plane of a Circular Aperture due to Incident Spherical Waves', *ibid.*, 1953, **24**, p. 110.
(79) HOGG, D. C.: 'The Electromagnetic Field near a Dielectric Lens', *ibid.*, 1954, **25**, p. 542.
(80) HUANG, C., KODIS, R. D., and LEVINE, H.: 'Diffraction by Apertures', *ibid.*, 1955, **26**, p. 151.
(81) HULL, G. F.: 'On the Use of Interferometer in the Study of Electric Waves', *Physical Review*, 1897, **5** (Series 1), p. 231.
(82) JENKINS, F. A., and WHITE, H. E.: 'Fundamentals of Physical Optics' (McGraw-Hill, New York, 1957).
(83) JONES, D. S.: 'Diffraction of a Waveguide of Finite Length', *Proceedings of the Cambridge Philosophical Society*, 1952, **48**, p. 118.
(84) JONES, E. M. T., and COHN, S. B.: 'Surface Matching of Dielectric Lenses', *Journal of Applied Physics*, 1955, **26**, p. 452.
(85) JONES, S. S. D., and BROWN, J.: 'Metallic Delay Lenses', *Nature*, 1948, **163**, p. 324.
(86) KAHAN, T.: 'Interferometer for Hertzian Microwaves', *Journal de Physique et le Radium*, 1947, **8**, p. 192.
(87) KAPRIELIAN, Z. A.: 'Dielectric Properties of a Lattice of Anisotropic Particles', *Journal of Applied Physics*, 1956, **27**, p. 24.
(88) KARP, S. N., and RADLOW, J.: 'On Resonance in Infinite Gratings of Cylinders', *Transactions of the Institute of Radio Engineers*, 1956, **AP-4**, No. 4, p. 654.
(89) KIRSCHBAUM, H. S., and CHEN, S.: 'A Method of Producing Broadband Circular Polarization Employing an Anisotropic Dielectric', *ibid.*, 1957, **MTT-5**, No. 3, p. 199.

(90) KLEIN, J. A., LOUBSER, J. H. N., NETHERCOT, A. H., and TOWNES, C. H.: 'Magnetron Harmonics at Millimetre Wavelengths', *Review of Scientific Instruments*, 1952, **23**, p. 78.
(91) KOCK, W. E.: 'Metallic Delay Lenses', *Bell System Technical Journal*, 1948, **27**, p. 58.
(92) KOCK, W. E.: 'Path-Length Microwave Lenses', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 852.
(93) KOCK, W. E.: 'Metal Lens Antennas', *ibid.*, 1946, **34**, p. 828.
(94) KODIS, R. D.: 'Diffraction Measurements at 1·25 Centimeters', *Journal of Applied Physics*, 1952, **23**, p. 249.
(95) KOTTLER, F.: *Annalen der Physik*, 1923, **70**, p. 405 and **71**, p. 457.
(96) LENGYEL, B. A.: 'A Michelson-Type Interferometer for Microwave Measurements', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 1242.
(97) LENGYEL, B. A.: Reflection and Transmission at the Surface of Metal-Plate Media', *Journal of Applied Physics*, 1951, **22**, p. 265.
(98) LEVINE, H., and SCHWINGER, J.: 'On the Theory of Electromagnetic-Wave Diffraction by an Aperture in an Infinite Plane Conducting Screen', *Communications on Pure and Applied Mathematics*, 1950, **3**, p. 355.
(99) LEWIN, L.: 'The Electrical Constants of a Material Loaded with Spherical Particles', *Journal I.E.E.*, 1947, **94**, Part III, p. 65.
100) MCFARLANE, G. G.: 'Surface Impedance of an Infinite Parallel-wire Grid at Oblique Angles of Incidence', *ibid.*, **93**, 1946, Part IIIA, p. 1523.
101) MARCUVITZ, N.: 'Waveguide Handbook' (McGraw-Hill, New York, 1951).
102) MARTIN, J. R., and SCHUNEMANN, C. F.: 'Measuring Wavelength in Millimetres', *Electronics*, 1953, **26**, p. 184.
103) MATHEWS, P. A., and CULLEN, A. L.: 'A Study of the Field Distribution at an Axial Focus of a Square Microwave Lens', *Proceedings I.E.E.*, 1956 Monograph No. 186 R, July (**103** C, p. 449).
104) MEECHAM, W. C., and PETERS, C. W.: 'Reflection of Plane-Polarized Electromagnetic Radiation from an Echelette Diffraction Grating', *Journal of Applied Physics*, 1957, **28**, p. 216.
105) MONTZER, J. R.: 'The Use of Dielectric Lenses in Reflection Measurements', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 252.
106) MEVEL, J.: 'Apparatus for the Experimental Study of the Diffraction of Centimetre Waves', *Journal de Physique et le Radium*, 1957, **18**, Supplement No. 3, p. 45A.
107) MEYER, E., and SEVERIN, H.: 'Absorption Devices for Centimetre Electromagnetic Waves and their Acoustic Analogues', *Zeitschrift für angewandte Physik*, 1956, **8**, p. 105.
108) MEYER, E., SEVERIN, H., and UMLAUFT, G.: 'Resonance Absorbers for Electromagnetic Waves', *Zeitschrift für Physik*, 1954, **138**, p. 465.
109) MEYER, E., KURTZE, G., SEVERIN, H., and TAMM, K.: 'A New Large Free Space Room for Acoustic and Short Electromagnetic Waves', *Acustica*, 1953, **3**, p. 409.
110) MIKAELYAN, A. L.: 'Methods for Calculating the Dielectric and Magnetic Permeabilities of Artificial Media', *Radioteckhnika*, 1955, **10**, p. 23.
111) MILLAR, R. F.: 'The Diffraction of an Electromagnetic Wave by a Circular Aperture', *Proceedings I.E.E.*, Monograph No. 196 R, September, 1957 (**104** C, p. 87).
112) MORITA, K., and COHN, S. B.: 'A Microwave Lens Matched by Simulated Quarter-Wave Transformers', *Transactions of the Institute of Radio Engineers*, 1956, **AP-4**, No. 1, p. 33.
113) MORSE, P. M., and RUBENSTEIN, P. J.: 'The Diffraction of Waves by Ribbons and by Slits', *Physical Review*, 1938, **54**, p. 895.
114) ORTUSI, J., and SIMON, J. C.: 'On the Focusing of a Wave', *Comptes Rendus Hebdomadaires des Séances de l'Académic des Sciences* (Paris), 1950, **230**, p. 521.
115) PRIMICH, R. I.: 'A General Experimental Method to Determine the Properties of Artificial Media at Centimetre Wavelengths Applied to an Array of Parallel Metallic Plates', *Proceedings I.E.E.*, Paper No. 1701 R, January, 1955 (**102** B, p. 26).
116) PRIMICH, R. I.: 'A Semi-Infinite Array of Parallel Metallic Plates of Finite Thickness for Microwave Systems', *Transactions of the Institute of Radio Engineers*, 1956, **MTT-4**, No. 3, p. 156.
117) PRIMICH, R. I.: 'Some Electromagnetic Transmission and Reflection Properties of a Strip Grating', *ibid.*, 1957, AP–5, No. 2, p. 176.
118) RAMSAY, J. F., and SNOOK, S. C. L.: 'Microwave Model Crystallography', *Electronic and Radio Engineer*, 1957, **34**, No. 5, p. 165.
119) RAMSAY, J. F.: 'Microwave Antenna and Waveguide Techniques before 1900', *Proceedings of the Institute of Radio Engineers*, 1958, **46**, p. 405.
120) LORD RAYLEIGH: 'On Pin-Hole Photography', *Philosophical Magazine*, 1891, **31**, p. 87.
121) ROBENHUSST, H.: 'Determination of Dielectric Properties from the Brewster Angle of Incidence and the Restored Azimuth at Wavelength below 30 cm', *Annalen der Physik*, 1955, **16**, p. 163.
122) ROBINSON, H. L.: 'Diffraction Patterns in Circular Apertures less than One Wavelength in Diameter', *Journal of Applied Physics*, 1953, **24**, p. 35.
123) RONCHI, L., and FRANCIA, G. T. DI.: 'An Application of Parageometrical Optics to the Design of a Microwave Mirror', *Transactions of the Institute of Radio Engineers*, 1958, **AP-6**, No. 1, p. 129.
124) ROW, R. V.: 'Microwave Diffraction Measurements in a Parallel Plate Region', *Journal of Applied Physics*, 1953, **24**, p. 1448.
125) RUST, N. M.: 'The Phase Correction of Horn Radiators', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 50.
126) RUZE, J., and YOUNG, M.: 'Experimental Determination of the Reflection Coefficient of Metal Plate Media', *Journal of Applied Physics*, 1951, **22**, p. 277.
(127) RYTOV, S. M.: 'Electromagnetic Properties of a Medium Comprising Thin Layers', *Zhurnal Eksperimentalnoĭ i Teoreticheskoĭ Fiziki*, 1955, **29**, p. 605.
(128) SCHAEFER, C., and GROSS, G.: 'Investigations Concerning Total Reflection', *Annalen der Physik*, 1910, **32**, p. 648.
(129) SCHMITT, H. J.: 'Wide-Band Resonance Absorber for Centimetre Electromagnetic Waves', *Zeitschrift für angewandte Physik*, 1956, **8**, p. 372.
(130) SEELEY, J. S.: 'A Spectrometer Method for Measuring the Electrical Constants of Lossy Materials', *Proceedings I.E.E.*, Monograph No. 248 R, July, 1957 (**105** C, p. 18).
(131) SEVERIN, H.: 'Surface Resistors for Centimetre Wave Technique', *Technische Mitteilungen PTT*, 1954, **32**, p. 209.
(132) SEVERIN, H.: 'Diffraction of Centimetre Electromagnetic Waves by Metal Disks', *Zeitschrift für angewandte Physik*, 1951, **3**.
(133) SEVERIN, H., and BAECKMANN, W. V.: 'Diffraction of Centimetre Electromagnetic Waves by Metal and Dielectric Disks', *ibid.*, 1951, **3**, p. 22.
(134) SEVERIN, H.: 'Application of the Methods of Optics to Calculations of Diffraction Phenomena in the Centimetre Wave Range', *Technische Mitteilungen PTT*, 1952, **30**, p. 347.
(135) SEVERIN, H.: 'Nonreflecting Absorbers for Microwave Radiation', *Transactions of the Institute of Radio Engineers*, 1956, AP–4, No. 3, p. 385.
(136) SHARPLESS, W. M.: 'Artificial Dielectrics for Microwaves', *Proceedings of the Institute of Radio Engineers*, 1951, **39**, p. 1389.
(137) SHINN, D. H.: 'Misfocusing and the Near-Field of Microwave Aerials', *Marconi Review*, 1956, **19**, No. 123, p. 141.
(138) SIMMONS, A. J., and EMERSON, W. H.: 'Anechoic-Chamber for Microwaves', *Electronic Industries and Tele-Tech*, 1953, July, p. 47.
(139) SIMON, J. C.: 'On the Focusing of a Wave', *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* (*Paris*), 1950, **230**, p. 1513.
(140) SNOW, O. J.: 'Transmission Characteristics of Inclined Wire Gratings', *Transactions of the Institute of Radio Engineers*, 1956, **AP-4**, No. 4, p. 650.
(141) SOLLOM, P. H., and BROWN, J.: 'A Centimetre-Wave Parallel-Plate Spectrometer', *Proceedings I.E.E.* Paper No. 2008 R, May, 1956 (**103**, B, p. 419).
(142) SOMMERFELD, A.: *Proceedings of the London Mathematical Society*, 1897, **28**, p. 395, and *Zeitschrift für Mathematik und Physik*, 1901, **46**, p. 11.
(143) STRATTON, J. A.: 'Electromagnetic Theory' (McGraw-Hill, New York, 1941).
(144) STUETZER, O. M.: 'Development of Artificial Microwave Optics in Germany', *Proceedings of the Institute of Radio Engineers*, 1950, **38**, p. 1053.
(145) SUSSKIND, C.: 'Obstacle-Type Artificial Dielectrics for Microwaves', *Journal of the British Institution of Radio Engineers*, 1952, **12**, p. 49.
(146) SWARUP, S.: 'Metal-Flake Artificial Dielectric', *Electronic and Radio Engineer*, 1958, **35**, p. 179.
(147) TALPEY, T. E.: 'Optical Methods for Measurement of Complex Dielectric and Magnetic Constants at Centimetre and Millimetre Wavelengths', *L'Onde Electrique*, 1953, **33**, p. 561.
(148) THEISSING, H. H., and MCGUE, J.: 'Wavelength Measurement in the Millimetre Region', *Review of Scientific Instruments*, 1955, **26**, p. 1203.
(149) TRENTINI, G. VON: 'Grids as Circuit Elements for Electromagnetic Waves in Space', *Zeitschrift für angewandte Physik*, 1953, **5**, p. 221
(150) TRENTINI, G. VON: 'Partially Reflecting Sheet Arrays', *Transactions of the Institute of Radio Engineers*, 1956, **AP-4**, No. 4, p. 66.
(151) TRENTINI, G. VON: 'Maximum Transmission of Electromagnetic Waves by a Pair of Wire Gratings', *Journal of the Optical Society of America*, 1955, **45**, p. 883.
(152) VAINSHTEIN, L. A.: 'The Diffraction of Electromagnetic Waves at a Grating Consisting of Parallel Conducting Strips', *Zhurnal Tekhnicheskoĭ Fiziki*, 1955, **25**, p. 847.
(153) VALENSI, G.: 'Polarization Filters and Polarization Correctors' *Annales de Télécommunication*, 1955, **19**, p. 230.
(154) VASSEUR, J. P.: 'Diffraction of Electromagnetic Waves by Apertures in Plane Conducting Screens', *L'Onde Electrique*, 1952, **32**, pp. 3, 55 and 97.
(155) WAIT, J. R.: 'The Impedance of a Wire Grid Parallel to a Dielectric Interface,' *Transactions of the Institute of Radio Engineers*, 1957, **MTT-5**, No. 2, p. 99.
(156) WALKER, G. B., and HYMAN, J. T.: 'The Use of Dielectric Materials to Enhance the Reflectivity of a Surface at Microwave Frequencies', *Proceedings I.E.E.*, Paper No. 2480 R, January, 1958 (**105**, B, p. 73).
(157) WARD, H. T., PURO, W. O., and BOWIE, D. M.: 'Artificial Dielectrics Utilising Cylindrical and Spherical Voids', *Proceedings of the Institute of Radio Engineers*, 1956, **44**, p. 171.
(158) WELLS, E. M.: 'Some Experiments on the Reflecting Properties of Metal-Tube Lens Medium', *Marconi Review*, 1954, **17**, p. 74.
(159) WHITEHEAD, E. A. N.: 'The Theory of Parallel-Plate Media for Microwave Lenses', *Proceedings I.E.E.*, Paper No. 1093 R, March, 1951 (**98**, Part III, p. 133).
(160) WOLF, E.: 'Microwave Optics', *Nature*, 1953, **172**, p. 615.
(161) WOOD, R. W.: 'Physical Optics' (Macmillan, London, 1934).
(162) WOONTON, G. A., CARRUTHERS, J. A., ELLIOTT, H. A., and RIGBY, E. C.: 'Diffraction Errors in an Optical Measurement at Radio Wavelengths', *Journal of Applied Physics*, 1951, **22**, p. 390.

621.396.677 : 621.396.933.2

The Institution of Electrical Engineers
Paper No. 2776 E
Mar. 1959
©

A NEW METHOD OF GENERATING A ROTATING RADIATION POLAR DIAGRAM

With Particular Reference to a Certain Navigational System

By H. W. HAWKES, B.Sc., Associate Member.

(*The paper was first received* 29*th November*, 1957, *in revised form* 10*th March, and in final form* 11*th September*, 1958.)

SUMMARY

In recent years, there has been a tendency to introduce new navigational aids which utilize rotating aerial arrays of large electrical dimensions in order to obtain highly accurate navigational information; one such example is the well-known system called Tacan. In general, the rotation of large arrays introduces severe mechanical problems which fix a lower limit of about 300 Mc/s to the operating radio frequencies of these aids. For many purposes it is desirable to operate in the 100 Mc/s region, and the paper describes a method by which this object may be achieved. Briefly, it consists of rotating an aerial array of small mechanical dimensions and allowing it to be coupled electrically to a static array of very large dimensions which acts as the final radiator of electromagnetic energy. The efficiency of the coupling and the form of the static array are such as to produce a final radiation which makes the rotatable array appear to have the large dimensions of the static array. In effect, the method is similar to the familiar 'goniometer' technique. The method is described with reference to its application in a new and more accurate form of Vor system (v.h.f. omni-range) known as Vorac. The paper also describes how the new method itself helps to improve the vertical elevational performance of a conventional Vor system.

(1) INTRODUCTION

(1.1) General

During the development of an improved form of v.h.f. omni-range system (Vor) known as Vorac (the letters 'ac' are an abbreviation for 'accurate'), it became obvious that the major obstacle to progress was the difficulty of generating the type of radiation polar diagram upon which the performance of the new system depended. The paper describes how, by using a new technique of generation, the difficulties were overcome. The new technique is reasonably general in nature and is not necessarily restricted to use in the Vorac system.

(1.2) The Rotating Radiation Polar Diagram

In order to clarify later description, it is necessary to give a brief outline of the principles involved in using rotating polar diagrams for the measurement of azimuth.

Fig. 1 shows a general aerial array which is uniformly rotated on a circle in the horizontal plane. The centre of rotation is at the point O, as shown. If the aerial array is being fed with electromagnetic energy, an electric field strength of

$$E = F_1(\phi)F_2(\phi) \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad (1)$$

can be received at a distant point P, where:

$F_1(\phi)$ is the expression for the amplitude variation of the polar diagram that is produced on turning the array.

$F_2(\phi)$ is an expression for the r.f. phase variation of the amplitude polar diagram, $F_1(\phi)$, about the point O.

ϕ is the angle between the line OP and a datum line joining O to a datum point on the array, as shown in Fig. 1.

ω is the angular frequency of the radiation field.



Fig. 1.—Rotation of the general array.

The composite polar diagram expressed by eqn. (1) can be rotated by rotating the aerial array of Fig. 1. If the speed of rotation is N revolutions per second, ϕ becomes a function of time, i.e.

$$\phi = 2\pi Nt + \phi_i$$

where ϕ_i is the value of ϕ at $t = 0$.

The aerial datum line can be arranged to pass through the northward meridian at $t = 0$, so that ϕ_i will represent the azimuth angle of the point P from that meridian. Thus the problem of azimuth determination resolves into one of finding the value of ϕ_i.

(1.2.1) The Use of the Amplitude Variation of the Polar Diagram.

If $F_2(\phi) = 1$, i.e. if there is no r.f. phase variation in the composite polar diagram, the rotating polar diagram effectively consists of an amplitude-modulated wave. The relationship of each component of the polar diagram $F_1(\phi)$ to each modulation component or signal will now be considered in order to show how ϕ_i can be determined.

It is convenient to express $F_1(\phi)$ as a Fourier series:

$$\begin{aligned} F_1(\phi) \sin \omega t = F_1(2\pi Nt + \phi_i) \sin \omega t = \\ [C_0 + C_1 \cos (2\pi Nt + \phi_i) + C_2 \cos 2(2\pi Nt + \phi_i) \\ + C_3 \cos 3(2\pi Nt + \phi_i) + \ldots \\ + C_n \cos n(2\pi Nt + \phi_i)] \sin \omega t \quad . \quad . \quad . \quad . \quad . \quad (2) \end{aligned}$$

where C_0 is an expression for the omni-directional component of the polar diagram which can be regarded as the carrier of an

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Hawkes is with Standard Telephones and Cables, Ltd.

amplitude-modulated wave; and $C_n \cos n(2\pi Nt + \phi_i)$ is an expression for the 'nth-order' sinusoidal variation of the polar-diagram amplitude, which can be regarded as the 'nth-order bearing' modulation of the a.m. wave, where n is a positive integer.

In all that follows, only an omni-directional, a first-order, and one nth-order ($n \neq 1$), polar diagram will be assumed to exist.

The values of ϕ_i and $n\phi_i$ can be determined by phase comparison of the two bearing modulations with first- and nth-order 'reference' signals of $\cos(2\pi Nt)$ and $\cos n(2\pi Nt)$ respectively. In practice, the reference signals are transmitted as additional modulations on the carrier C_0. The exact details of the reference generation, modulation, detection at the point P and the phase comparison will not be considered here; however, details for some typical practical systems have been given.[1,2,3,4]

The numerical value of ϕ_i, derived from the phase comparison of the first-order bearing and reference signals, is presented on a phase meter with a range of 0–360°; thus the azimuth angle is directly indicated on that meter. Equivalent phase comparison and presentation of the nth-order signals will result in a meter indication that is representative of any one of n possible azimuth positions of the point P, each position being separated by an angle of $360/n$ degrees from its immediate neighbour. This 'ambiguous' measurement has the advantage of improving the accuracy of azimuth indication by a factor n. The resultant ambiguities are resolved by the phase comparison of the first-order signals. Details of the ambiguity resolution and the reasons for the improved accuracy can be found in References 3–6.

Up to recent years, only the first-order signal measurement of azimuth has been used in practical navigational aids, an example of such an aid being Vor.[7,8] Recent suggestions[6] have led to the construction of aids utilizing one ambiguous measurement of ϕ_i, for example Tacan[4] and Vorac.[9]

(1.2.2) **Brief Descriptions of Tacan, Vor and Vorac Systems.**

Tacan uses first-order modulation signals of 15 c/s and 'ninth-order' modulation signals of 15 × 9 c/s. An aerial array, producing a first-order and ninth-order polar diagram, is spun at a speed of 15 c/s to give the bearing signals. The aerial array producing the omni-directional component of the polar diagram is not rotated. The operating radio frequency of the system is about 1 000 Mc/s. Tacan will be coming into extensive use in the near future.

Vor uses first-order modulation signals of 30 c/s produced by the rotation of a first-order polar diagram at 30 c/s (again the omni-directional polar diagram is not rotated). The operating radio frequency is about 100 Mc/s, and the system is in widespread use at the present time.

One type of Vorac uses the first-order modulation signals of Vor plus fifth-order modulation signals. The system differs from Tacan in two vital respects:

(*a*) The operating radio frequency is that of Vor (i.e. 100 Mc/s).
(*b*) Each polar diagram is generated by a separate aerial so that the fifth-order polar diagram may be rotated at a speed that is one-fifth that of the first-order polar diagram. Thus the two amplitude-modulation signals produced by the polar diagrams become $\cos n(2\pi Nt/n + \phi_i)$ and $\cos(2\pi Nt + \phi_i)$, and it can be seen that only one reference signal of $\cos 2\pi Nt$ is necessary for the determination of ϕ_i and $n\phi_i$. Further details of this procedure and of certain modifications are given in References 1 and 2.

The Vorac system uses the new method of polar-diagram generation and at present is an experimental system.

The above information will provide the background of the operational use to which the methods described in the paper are directly applicable. However, mention must be made of other navigational uses of rotating polar diagrams which could use the new method.

(1.2.3) **Other Navigational Uses of Rotating Polar Diagrams.**

It is possible to produce a polar diagram having a phase variation of $F_2(\phi)$ and use this to provide means for the measurement of ϕ_i.[10] The new method of polar-diagram generation has been used in a phase-modulation system of this type.

It is also possible to use combinations of the functions $F_1(\phi)$ and $F_2(\phi)$ to measure the angle ϕ_i; however, this type of system has not been developed to any considerable extent.

Radar navigational systems use partial or continuous rotation of polar diagrams to a considerable extent, and it may be possible to apply the present work to some of these systems.

(2) THE PROBLEM OF GENERATING MULTI-LOBE POLAR DIAGRAMS

(2.1) General

An nth-order polar diagram can be produced from the nth-order aerial array ($n = 5$) shown in Fig. 2. The array consists of a number of $\frac{1}{2}\lambda$ vertical dipoles uniformly spaced on a circle

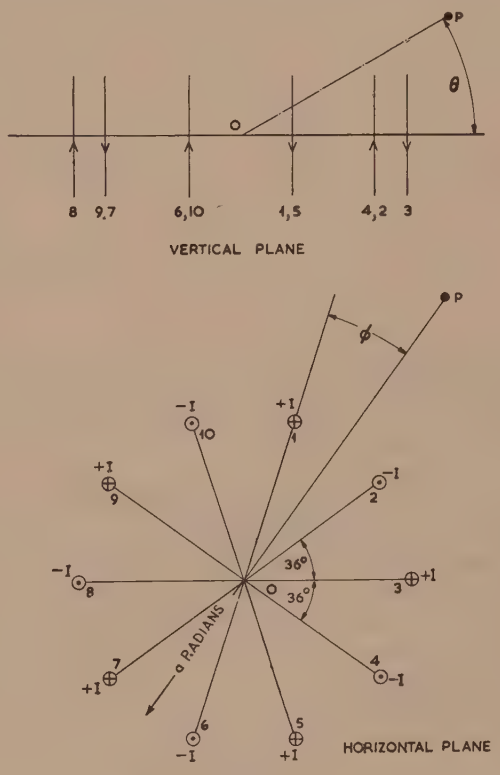


Fig. 2.—Simple nth-order array.

of radius a radians. A 180° phase difference in current exists between adjacent aerials; thus, each aerial has a current of either I or $-I$ amperes as is shown in the diagram. The aerial array is analysed mathematically in Section 12.2.4, where it is shown that the distant field strength, at P, in the horizontal plane ($\theta = 0$) of the array can be written as

$$k_x I J_n(a) \cos n\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $2n$ is the total number of aerials in the array shown and k_x is a radiation constant.

If eqn. (3) is plotted in polar form, it will be seen that the diagram consists of $2n$ lobes, each having a polarity, or r.f. phase, opposite to that of an adjacent lobe. This polar diagram is often known as an nth-order multi-lobe polar diagram. A typical fifth-order multi-lobe polar diagram is shown in Fig. 3.

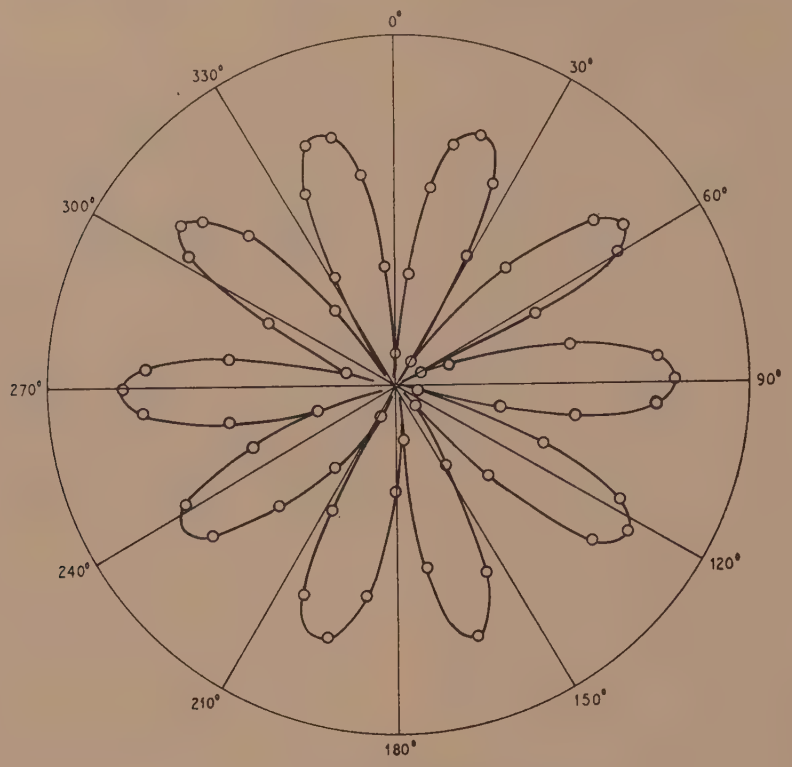


Fig. 3.—Typical fifth-order polar diagram.
○ Measured amplitude points.

An inspection of eqn. (3) shows that the magnitude of distant field strength is dependent on the radius of the array, because of the term $J_n(a)$. This dependence leads to certain problems that will now be considered.

(2.2) Effect of the Array Radius on Radiation Efficiency and Aerial Impedance

(2.2.1) Efficiency and Impedance Factors.

The input impedance of a single radiating element or aerial (i.e. an electric dipole) can be divided into the following:

(*a*) The radiation resistance R.
(*b*) The loss resistance r.
(*c*) The reactance X.

Parts (*a*) and (*b*) determine the radiation efficiency of the element, i.e.

$$\eta = \frac{R}{R+r} \times 100\% \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The absolute value of aerial impedance is relatively unimportant, since an impedance transformation can usually be employed to give satisfactory matching to standard feed cables. The most important factor in the aerial impedance is its rate of change with respect to the frequency of the applied radio signal.

In the paper it has been found convenient to define this rate of change in terms of the familiar *LC* circuit concept of Q-factor; i.e.

$$Q = \frac{f}{(f_1 - f_2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where f is a frequency at which the circuit input impedance is purely resistive—a condition obtained by reactance compensation if necessary; f_1 and f_2 are the frequencies at which the phase angle of the input impedance has increased to $+45°$ and $-45°$, the resistive component of the impedance still remaining at the value of R, and $f_1 > f > f_2$.

Let ΔX, the change in reactance for every cycle-per-second deviation from the frequency f, be constant over the range f_1 to f_2. The tangents of the phase angles (i.e. $45°$ and $-45°$) are given by $[(f_1 - f)\Delta X]/R$ and $[(f_2 - f)\Delta X]/R$ at f_1 and f_2 respectively, from which it can be seen that $f_1 - f_2 = 2R/\Delta X$.

Substituting in eqn. (5), the Q-factor is given by

$$Q = \frac{f\Delta X}{2R} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Eqn. (6) can be used to define the 'radiation' Q-factor of an aerial input impedance at the one given frequency f, without reference to f_1 and f_2, i.e. the Q-factor, the radiation resistance R and ΔX may have different values at f_1, f_2 and f. This definition is used throughout the paper, as the aerial arrays considered have a pronounced variation of R with frequency which causes the Q-factor to change with frequency.

(2.2.2) Variation of Radiation Resistance and Reactance.

An aerial in an nth-order array of the type shown in Fig. 2 will have its input impedance affected by the electrical distance between aerials (and hence by the array radius) because of mutual coupling. From the geometry of the array and known mutual impedance tables, it is possible to calculate the effect. The calculations are rather tedious, but, fortunately, as far as the radiation resistance component is concerned, it is only necessary to calculate the resistance at any one given radius, as values at any other radius can then be found by the following simple method.

Let R_{max} be the radiation resistance of one aerial in any array of the type shown in Fig. 2. The radius of this array is

a_{max} radians, the value at which $J_5(a)$ reaches its maximum. The current in each aerial is I_{max} and is produced by the supply of an r.f. power P_T.

Let R be the radiation resistance of one aerial in another array of the type shown in Fig. 2. The radius of this array is taken as a radians, and the current in each aerial is I produced by the supply power P_T.

The radiation efficiency of both arrays is assumed to be 100%. It will also be assumed that the shapes of the polar diagrams of the arrays are identical—as implied by eqn. (3)—so that the same radiation field should be produced by both, i.e. the field strength at the point P for either array is

$$kIJ_n(a) \cos n\phi = kI_{max}J_n(a_{max}) \cos n\phi$$

and also the power fed into each aerial is

$$P_T = I^2R = I_{max}^2R_{max}$$

From the two equations, it is seen that

$$R = R_{max}\left[\frac{J_n(a)}{J_n(a_{max})}\right]^2 \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The above assumption of equal polar-diagram shapes is strictly applicable only to the horizontal plane, because there is a change of shape with radius in the vertical plane [see eqn. (14)]. The change is not great for $a < a_{max}$, and, for the present purpose, the assumption can be regarded as valid.

Thus, if the radiation resistance and efficiency of a given aerial when radiating by itself and the value of R_{max} produced by the given aerial in an array of the type shown in Fig. 2 are known, it is possible, by means of eqns. (4) and (7), to calculate the radiation resistance and efficiency (in future to be termed the array-radiation resistance and efficiency) of the given aerial when it is acting as a member of an array whose radius is less than a_{max} radians. Calculation of the array-radiation Q-factor by eqn. (6) will require knowledge of ΔX. ΔX, of a single aerial radiating alone, can be found from an aerial reactance/frequency curve, which in turn can be estimated from experimental evidence or from the extensive data that are supplied by radio reference handbooks for the usual shapes and sizes of aerials. These estimated values of ΔX will change when the aerial is operating as a member of an nth-order array, but the magnitude of the change may be found from mutual impedance considerations; however, these considerations show that the change is small and may be neglected as far as the present work is concerned.

The procedure given above will now be used in order to make a practical study of the problems involved in the construction of nth-order array.

(2.2.3) Limitations on the Array Radius

The following calculations are based on an array of the type of Fig. 2 that is required to radiate a fifth-order polar diagram at a Vor frequency of 114 Mc/s.

It is reasonably simple to construct v.h.f. $\frac{1}{2}\lambda$ dipoles with radiation efficiencies and Q-factors of the order of 99% and unity respectively. They could be used in an array of about 6 rad radius where they would have individual maximum array-radiation resistances of R_{max}, which would be about double the normal radiation resistance of one dipole radiating alone—a fact ascertained from mutual impedance considerations. By eqns. (4) and (6), this would produce array efficiencies and Q-factors of 99·5% and 0·5.

This array is ideal from the electrical view-point. However, an electrical radius of 6 rad is an array diameter of about 16 ft when the operating frequency is 114 Mc/s, and rotation of such a large array at speeds above 150 r.p.m. would be extremely difficult. An array of 6 ft diameter would be fairly convenient to spin, but the array-radiation resistance would fall to $0{\cdot}001R_{max}$ [from eqn. (7)]. The array-radiation Q-factors and efficiencies would thus become 500 and 16%. A good idea of the general difficulties of small-radius operation can be obtained by referring to the graph of Fig. 4, which shows the steep fall in array-

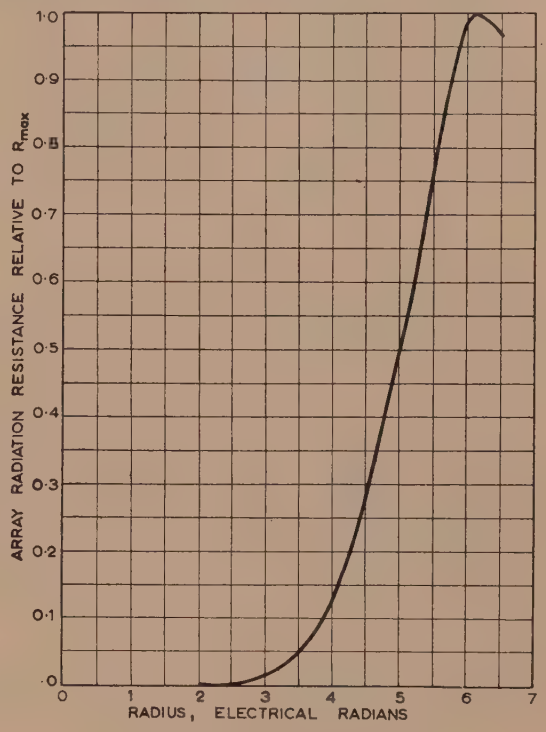


Fig. 4.—Variation of array-radiation resistance of fifth-order array.

radiation resistance, relative to R_{max}, with decreasing radius. The graph was calculated from eqn. (7). In practice, operation of an nth-order array of small radius will result in the radiation of a polar diagram of quite random shape due to the following factors:

(i) It is difficult to match each high-Q aerial to practical transmission lines and still obtain a given amplitude and phase of current in each aerial.

(ii) Slight variations in small aerial spacings will produce large variations in each aerial input impedance and so accentuate the difficulty in (i).

Such difficulties could possibly be overcome, but it is clear that the amount of work involved would be prohibitive, especially as large reductions in array radius would not only accentuate the above factors but would soon reduce the array-radiation efficiency to unworkable levels. A practical study of such factors led to the conclusion that it was not advisable to reduce the radius much below 3·7 rad (i.e. a 10 ft diameter at 114 Mc/s) if the array was to be manufactured using normal production facilities and techniques. Correct operation of the array was in fact obtained down to about 3·5 rad, but this was only by using special construction, adjustments and testing that would not normally be tolerated.

A few solutions to the general problem were then considered:

(*a*) To rotate the polar diagram by an electronic method of rotation as in certain forms of normal Vor equipment. This solution is not very practical because of the complexity of the resultant aerial and r.f. feed system.

(*b*) The use of $2n$ individual array elements of much greater

directivity than electric dipoles will give improved radiation characteristics at reduced radii. The disadvantage is that these elements would have to be electrically much larger than the dipole, and so would be bulky for spinning purposes, and difficult to space around the small circumference of the array.

(*c*) To reduce the rotational speed to about 15–30 r.p.m. The disadvantage is that these low speeds would mean fifth-order bearing signals of the order of a few cycles per second. It is generally accepted that the 15 c/s bearing signal of Tacan is about the lowest frequency that could be employed for reliable azimuth determination.

It would thus appear that higher *n*th-order operation—for the difficulties increase with the higher values of n since a_{max} increases with n—is only possible at higher radio frequencies of approximately 300 Mc/s and upwards. The desirability of operation in the Vor frequency band, and also the attraction of drastically reducing the diameter of the rotating array, led to the consideration of a new method of generating the required polar diagram.

(3) A NEW METHOD OF POLAR-DIAGRAM GENERATION

The new method of polar-diagram generation does not reduce the diameter of the array, but it does reduce that of the mechanically rotating part.

A rotatable array of small diameter, theoretically capable of producing an *n*th-order polar diagram, is electrically coupled to a static array of very large diameter which acts as the final radiator of electromagnetic energy. The efficiency of coupling and the form of the static array are such as to produce a rotating polar diagram defined mainly by the small array, but giving that array the radiation efficiency and Q-characteristics that would be expected from an *n*th-order array of the same diameter as the static array. One arrangement that uses this procedure is shown in Fig. 5.

A metal cylinder, of radius a_s, and 3·28 rad in height (shown as detail 9), is slotted in ten places, and each slot is fed with a given amplitude of voltage which is 180° out of phase with the voltage on adjacent slots; thus, theoretically, the array would be capable of producing a horizontally polarized fifth-order polar diagram (Section 12.3). The numerical value of a_s in Fig. 5 is arranged to be about 0·9 rad, and so this array, when radiating alone, could be expected to have a radiation Q-factor and efficiency similar to those produced by an array of the type shown in Fig. 2 (i.e. about 10^6 and 0·0001%). However, the slotted cylinder is not alone, but is surrounded by the collection of static waveguides shown in Fig. 5. One waveguide is illustrated by the numerals 1, 2, 3, 4 and 5, 6, 7, 8, the lower and upper sets defining the input and output ends of the guide. The input end of the guide 'samples' the field of the array, at a radius of $a_s + \Delta a$ and along the height of the array, so that waveguide operation is in the TE_{10} mode. This field is transferred to the guide outputs which act as final radiators of the array at a radius a. In future it will be convenient to call the central array the 'exciter array', and the waveguide system, the 'aerial transducer' or more briefly 'transducer'. The radiation characteristics of the rotatable exciter array must be similar to those obtained from an array of radius a, since the rotating field at the radius $a_s + \Delta a$ is transferred to produce the same rotating field at the larger a. An analysis of this effect is given in Section 12.4, where it is shown how the radiation resistance of the whole arrangement in Fig. 5 is dependent on the parameters a_s and Δa. For $\Delta a = 0{\cdot}04$ rad and $a_s = 0{\cdot}9$ rad, it is possible to obtain a radiation resistance from the whole arrangement that is 0·6 of that obtained from a slotted cylinder of radius a. Thus a reliable fifth-order polar diagram can be produced from the whole arrangement if a is greater than about 3·7 rad. It should be noted that the size of the exciter array (i.e. a radius of 16 in at 114 Mc/s) is well inside the limit set by the mechanical consideration of rotation (Section 2.2.3).

A further description of the new method of generation is given in Reference 11.

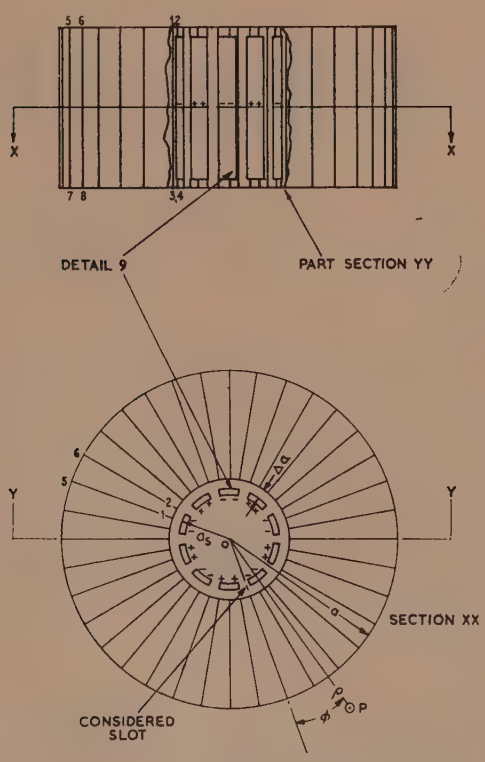


Fig. 5.—Waveguide transducer.

(3.1) Experimental Verification of the Transducer Principle

The transducer principle was investigated experimentally by work on a model designed for the band 300–370 Mc/s.

The electrical dimensions and other parameters of the model at 312 Mc/s, were as follows:

Height of exciter array, $\frac{1}{2}\lambda$; n, 5; a, 3·7 rad; a_s, 0·9 rad; Δa, 0·04 rad.

The theoretical radiation resistance of the transducer arrangement relative to that obtained from a slotted cylinder of radius a is 0·6. Thirty-six waveguides were used to sample the exciter array field, each having a cut-off frequency of 300 Mc/s.

The purpose of the first test was to show that the exciter array was incapable of giving a usable fifth-order polar diagram when radiating alone; in practice, a completely random polar diagram was obtained and the amount of radiation was extremely small. On rotating the energized exciter array in the transducer, it was found that a good fifth-order polar diagram was obtained; in addition, field strength measurements indicated that the radiation efficiency was satisfactory. A typical polar diagram obtained from a test of this nature is shown in Fig. 3.

The achievement of the characteristic fifth-order polar diagram was regarded as encouraging, but further information was required to show that radiation was definitely taking place from the outside circumference of the transducer, and that it was possible to obtain impedance characteristics from the transducer and exciter arrangement which would be approximately equivalent to those obtainable from a slotted cylinder array of 3·7 rad radius.

Two experiments were made to obtain the information, namely:

(*a*) A simple r.f. detector unit was used to measure the field strength very close to the outside of the transducer. The result of the measurement showed that there were ten small regions of high field strength similar to those found on a metal cylinder with ten energized slots.

(*b*) The impedance at the junction point of the 10 feeds to the exciter slots was measured over the stated frequency range. The array radiation Q-factor was then found from these data. All simple types of radiator seem to give about the same result when connected as members of a multi-lobe aerial array. It was thus reasoned that the theoretical radiation Q-factors of an array of the type shown in Fig. 2 should be similar to those of the slotted-cylinder type. By using the curve shown in Fig. 4 and assuming the slots to have radiation Q-factors of unity, it was thus possible to calculate approximate values of the Q-factor for the slotted-cylinder type of array, and hence find the array radiation Q-factors for the transducer by taking into consideration the 0·6 reduction of radiation resistance.

Fig. 6 shows the experimental and calculated values of array radiation Q-factor plotted against the outside radius a of the transducer; the close agreement between the curves indicates how the electrical performance of the rotatable array has been increased to that obtainable from an array of radius a.

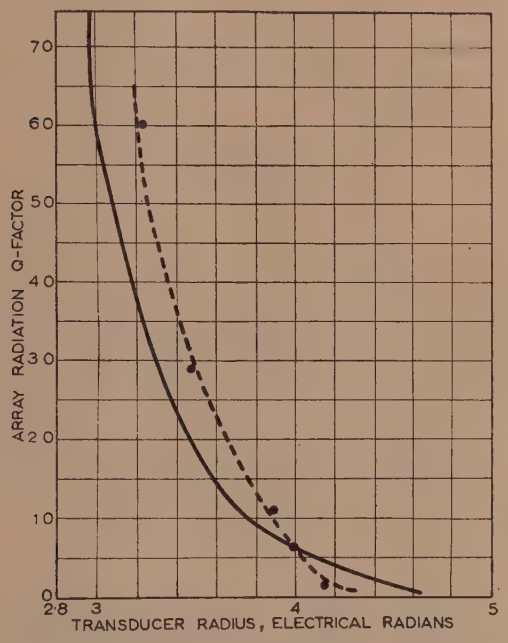


Fig. 6.—Variation of array-radiation Q-factor.
—— Calculated.
---- Experimental.

(4) THE TRANSMISSION-LINE TRANSDUCER

The experiments on the waveguide transducer were intended to provide design data for a v.h.f. array suitable for a Vorac type of navigational aid. However, during the experiments it was realized that the waveguide transducer would be extremely heavy when scaled up in dimensions. Further experiments were then carried out, in which light-weight transmission lines were used to replace the waveguide walls: Fig. 7 shows a 'transmission-line' transducer constructed on this basis. The results on this transducer were comparable with those on the original waveguide transducer, and all transducers for Vorac systems now use this form of construction.

(5) DEVELOPMENT AND FURTHER APPLICATIONS OF THE TRANSDUCER PRINCIPLE

(5.1) Operation in Any Given Frequency Band

The transducer of Section 3.1 operates in the band of 300–370 Mc/s. Further experimental work in the v.h.f., u.h.f., and

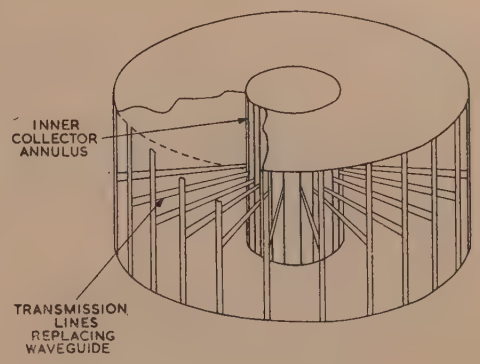


Fig. 7.—Transmission-line transducer.

even s.h.f. bands has shown that the principles are applicable over a wide range of frequencies. At high frequencies it is advisable to increase the electrical width of the annular gap Δa so that the mechanical gap never falls much below $\frac{1}{4}$ in; otherwise the conversion from one frequency band to another is a simple one of scaling up or down.

(5.2) Operation with Vertical Polarization

Previous description has been concerned with transducers radiating a horizontally polarized electric field. A form of line transducer that is able to radiate a vertically polarized electric field is shown in Fig. 8. Transmission lines connect the outer

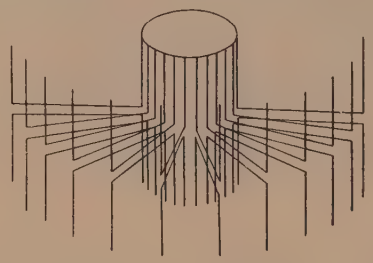


Fig. 8.—Sector of a vertical polarization transducer.

ring of $\frac{1}{2}\lambda$ dipoles to an inner ring of $\frac{1}{2}\lambda$ dipoles in which is placed an exciter array of $\frac{1}{2}\lambda$ dipoles—e.g. an nth-order array of the type shown in Fig. 2. If the annular spacing between the inner ring and exciter array is very small, it is apparent that the transducer action can take place as for the transmission-line transducer of Section 4. Radiation, arising from the transmission lines, and even directly from the exciter array, can prove troublesome, but one convenient method of overcoming this difficulty is shown

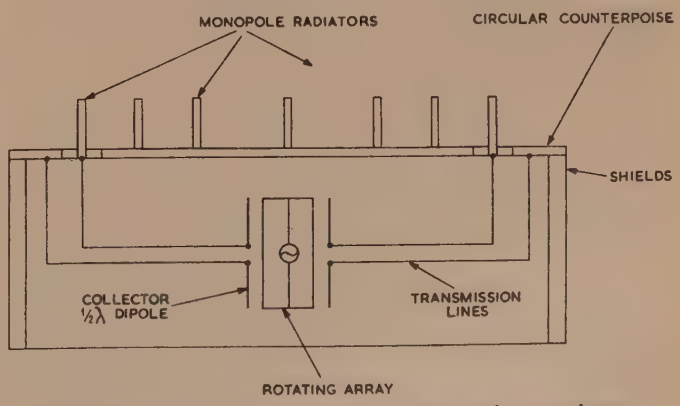


Fig. 9.—Diametric section through counterpoise transducer.

in Fig. 9. A study of this diagram shows that the outer ring of static radiators are monopoles mounted on a counterpoise, but the inner collecting ring and exciter elements are $\frac{1}{2}\lambda$ dipoles. These latter elements, together with the transmission lines, are under the counterpoise and cannot radiate. Such a transducer has been constructed and has given satisfactory results.

(5.3) Operation with Reduced Spinning Diameters

The diameter of the exciter array, described in Section 3, was originally judged to be highly convenient for spinning. Further reduction in spinning diameters was not attempted, although later experimental work indicated that an efficient fifth-order transducer polar diagram was possible at an exciter array radius of 0·8 rad. At the time of writing, no application is foreseen of either lower-frequency operation or higher spinning speeds, so that this aspect has received no further consideration.

(5.4) Reducing the Height of the Exciter Array

The exciter array has been quoted to be of 3·14 rad ($\frac{1}{2}\lambda$) height. It is possible to reduce this height and still obtain reasonably efficient radiation; array heights of one radian (approximately $\frac{1}{6}\lambda$) were found to give reasonable multi-lobe polar diagrams. A complete set of results is not available, but it may be said that the lower limit to height of the exciter array appears to be in the region of $\frac{1}{10}\lambda$. The limit is thought to be due to the increased coupling factors between array slots or dipoles and a reduced coupling between the array and inner ring of collecting slots or dipoles.

(5.5) Other Possible Uses of the Transducer

The paper is concerned with the use of the transducer as part of a multi-lobe Vor equipment, but Tacan, as a multi-lobe type of beacon, could use the new method of pattern generation. However, owing to the high operating frequency, the problem of spinning an array of large electrical radius is not of fundamental importance.

By using a transducer containing an exciter array of only one energized aerial, it is possible to rotate an energized source of radiation around the circumference of the transducer and so help to form a phase-modulation type of navigational system. An experimental equipment of this type has been constructed.

One other rather important application of the transducer is in the provision of a first-order bearing signal in a multi-lobe Vor system; this aspect will now be considered.

(6) THE USE OF THE TRANSDUCER FOR THE GENERATION OF THE OMNI-DIRECTIONAL AND FIRST-ORDER POLAR DIAGRAMS

As already indicated, the use of a first-order polar diagram is necessary in order to resolve the ambiguities of an nth-order multi-lobe navigation aid. This requirement can be met by the use of the aerial arrangement shown in Fig. 10.

Detail 3 is a horizontal radiating dipole (of 'fan' or 'bow-tie' form) capable of being mechanically rotated in order to provide the first-order bearing signal. Detail 2 is a horizontal loop that is fed in a number of places so as to achieve constant current in the loop; the radiation pattern of this loop is, of course, omni-directional and provides a carrier r.f. signal. The currents in the loop and dipole have identical radio frequencies, but their relative r.f. phase is such as to cause a maximum amplitude modulation of the carrier by the bearing signal (Section 12.4.3). Detail 1 is a metal cylindrical cavity whose walls are slotted as shown. The slots, which are approximately $\frac{1}{2}\lambda$ in height, allow the radiations from the dipole and loop aerials to escape into free space and thus, in effect, act as the final radiators of the

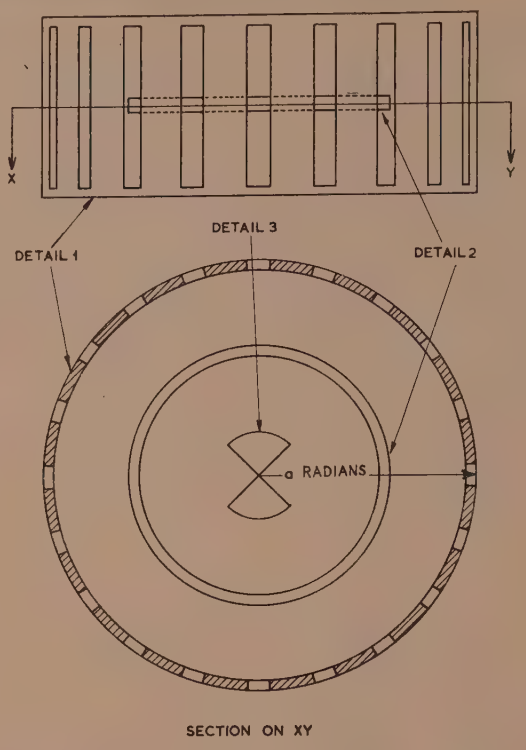


Fig. 10.—Transducer cavity.

complete arrangement. It will be recognized that the cavity can be regarded as a form of transducer in which the sampled field is only transmitted over the thickness of the cavity walls. The purpose of this transducer cavity is not to reduce the electrical dimensions of the rotating dipole but rather to make use of the fact that all final radiation must take place from the outside surface of a transducer, even if the coupling between the exciter array and that surface is very small (see Section 12.4.3). The advantage obtained from this property will be seen in the next Section.

(7) THE EFFECT OF THE TRANSDUCER ON THE VERTICAL COVERAGE OF VOR SYSTEMS

In Vor systems, it is usually found that the amplitude modulation of the bearing signal will fall with increase in the angle of elevation, θ, of the distance reception point P—two such typical points are shown in Figs. 2 and 13. The reason for the fall is that the amplitudes of the first- and nth-order polar diagrams (providing the modulations) will change in relation to that of the omni-directional polar diagram (providing the carrier). The elevation angle at which the modulation falls to one-half of that existing at the ground plane ($\theta = 0$) is termed the 'vertical-coverage' angle of the system. It is a Vor operational requirement that this angle shall be as high as possible. Section 12.4.3 deals with the analysis of the modulation fall, and so enables the vertical-coverage angles of a Vorac system to be compared with the angle known to exist for a normal Vor system.

(7.1) Vertical Coverage of the First-Order Bearing Signal in a Vorac System

From eqns. (30) and (31) it is possible to find the fall in modulation coefficient, relative to that existing at ground level, for the transducer cavity of the type shown in Fig. 10. The fall is plotted in Fig. 11 for a transducer of outside radius 3·7 electrical radians. It should be noted that the modulation coefficient

approximately constant up to 60° elevation, and then falls to ve a vertical-coverage angle of about 82°.

The vertical-coverage angle of a normal Vor system is only out 60°, since the aerial arrays used to generate the required olar diagrams are usually of a small array diameter, and this roduces a fall of modulation with elevation proportional to os θ. This cosine law can be shown to apply when the diameter the cavity shown in Fig. 10 is very small, i.e. when 'a' tends zero. The typical cos θ law obtained from such an analysis also shown in Fig. 11. The improvement, over the cos θ law,

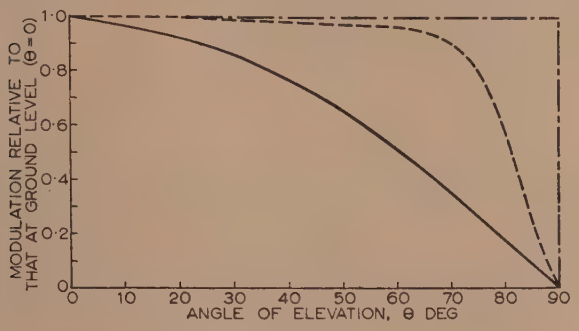


Fig. 11.—Fall of relative modulation coefficient.
—— Small-aperture cos θ system ($a \to 0$).
---- $a = 3{\cdot}7$ rad.
—·— $a = \infty$.

creases with the radius a; at an infinite radius, the overhead overage is complete, as is shown in Fig. 11. The improved overage is a direct result of using a large-diameter transducer. he improvement is, of course, also applicable to normal Vor ystems if a large-diameter transducer is used to radiate the ppropriate polar diagrams. The increase of vertical coverage 82° is regarded by Vor operators as a very desirable haracteristic.

(7.2) Vertical Coverage of the *n*th-Order Bearing Signal in a Vorac System

The vertical coverage of the nth order bearing signal can also be ound from an analysis similar to that given in Section 10.4.3. rom this analysis, it can then be shown that carrier and fifth-rder signals, radiated from a transducer of 3·7 rad radius, will ve a vertical-coverage angle of 30°. Thus all the vertical-overage requirements above 30° must be met by the first-order nbiguity resolver. The angle can only be increased beyond 0° by increasing the outside radius of the transducer.

) ARRANGEMENT OF OMNI-DIRECTIONAL, FIRST- AND *n*th-ORDER ARRAYS IN A TYPICAL MULTI-LOBE SYSTEM

Fig. 12 is a photograph of a Vorac aerial array. B and D dicate the nth-order and cavity transducers respectively. Only he omni-directional aerial array is used, located in the cavity as ustrated in Fig. 10.

The design of the nth-order transducer in the arrangement own has been based on the models already described, so that e common radius of the transducers is 3·7 rad (i.e. 10 ft at 4 Mc/s), which produces a good fifth-order polar diagram and creases the vertical-coverage angle of the first-order signals to °. A complete Vorac system, using the aerial arrangement, s been constructed and tested and has given very encouraging sults.[12]

(9) CONCLUSION

The paper has introduced a new method of generating rotating lar diagrams and has shown how the method can be used to overcome certain problems and defects that are apparent in some of the latest types of navigational aids.

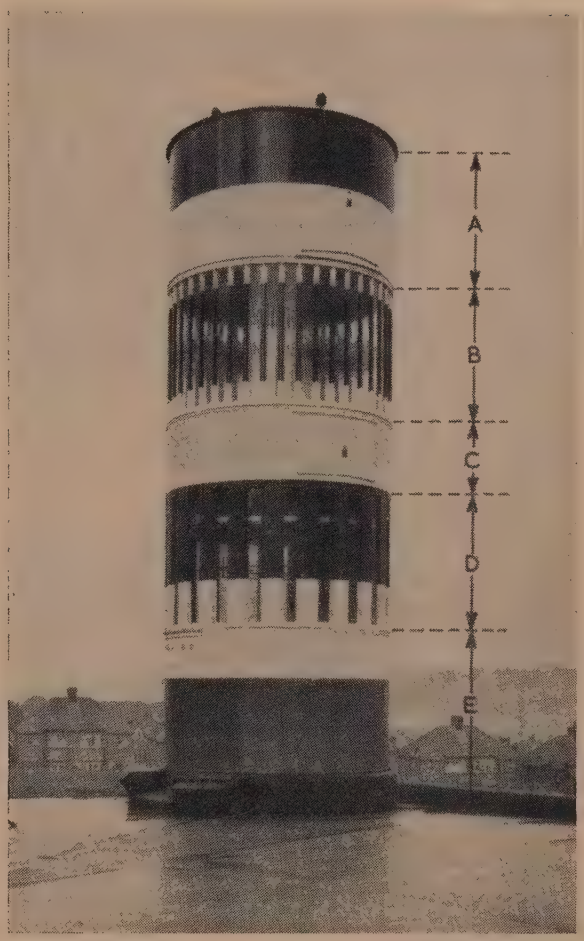


Fig. 12.—Transducer arrangement in Vorac.

The method has been shown to be practical and has also been analysed theoretically when it is operated under certain ideal conditions. A complete and rigid analysis for the method, when used in the form of a practical apparatus, has not yet been developed; however, the analysis, as given, seems to agree reasonably well with the experimental results obtained from one practical form of apparatus.

(10) ACKNOWLEDGMENTS

Acknowledgments are due to Standard Telephones and Cables Ltd., for permission to publish the paper. The author wishes to express his thanks to Mr. C. W. Earp for his helpful advice and discussion.

(11) REFERENCES

(1) HAWKES, H. W.: British Patent No. 802258.
(2) STRONG, C. E.: 'General Aspects of Short-Range Rho-Theta Systems', *Proceedings I.E.E.*, Paper No. 2578 R, March, 1958 (**105** B, Suppl. 9, p. 284).
(3) DODINGTON, S. H.: 'Airborne Tacan Equipment', *Electrical Communication*, 1956, **33**, p. 60.
(4) COLIN, R. I., and DODINGTON, S. H.: 'Principles of Tacan', *ibid.*, 1956, **33**, p. 11.
(5) MASONSON, M.: 'Error Reduction in Tacan Bearing-Indicator Facility', *ibid.*, 1956, **33**, p. 85.

(6) CLEAVER, R. F.: British Patent No. 635701.
(7) LUCK, D. G.: 'An Omni-directional Radio Range', *Radio Corporation of America Reviews*, 1941, **6**, p. 55; 1942, **6**, p. 344; 1946, **7**, p. 94.
(8) HARLEY, H. C., ANDERSON, S. R., and KEARY, H. F.: 'The C.A.A. V.H.F. Omni-Range', *Proceedings of The Institute of Radio Engineers*, 1951, **39**, p. 1506.
(9) 'Vorac High-Accuracy Vor System', *Civil Aviation Radio News Letter*, 1956, No. 24, p. 14.
(10) EARP, C. W., and COOPER-JONES, D. L.: 'The Practical Evolution of the Commutated-Aerial Direction-Finding System', *Proceedings I.E.E.*, Paper No. 2569 R, March, 1958 (**105** B, Suppl. 9, p. 317).
(11) HAWKES, H. W.: British Patent No. 744615.
(12) 'Trials of the Vorac System', *Civil Aviation News Letter*, 1957, No. 26, p. 18.
(13) JORDAN, E. C.: 'Electromagnetic Waves and Radiating Systems' (Prentice Hall, 1950), p. 593.

(12) APPENDICES

(12.1) A Discussion on Zero- and *n*th- Order Polar Diagrams

The Appendices are concerned with an analysis of certain types of aerial arrays suitable for producing zero- (i.e. omni-directional) and *n*th-order (including first-order) polar diagrams.

The polar diagrams can be produced by a number of aerials that are uniformly spaced on the circumference of a circle. The method of analysing such an array is to refer all radiation from the individual aerials to a common centre of symmetry and then to expand the resultant expression in series form.

(12.2) Polar Diagrams produced by Vertical Electric and Magnetic $\frac{1}{2}\lambda$ Dipoles

(12.2.1) Polar Diagram of a Single Dipole.

Consider the single dipole No. 1 at a distance of a radians from the common centre O, as is shown in Fig. 2, and consider a distant point P ($OP \gg 10a$), at an angle ϕ from the line between dipole No. 1 and the point O, and at an elevation angle θ from the horizontal ground plane of $\theta = 0°$.

The amplitude polar diagram of the dipole is omni-directional in the horizontal plane, and of $\cos\theta$ form in the vertical plane. If the dipole is rotated about O on the circumference of $2\pi a$, it is seen that the field strength at P will be phase modulated by an amount $a\cos\theta\cos\phi$ radians. Thus if the dipole is fed with a current I, the field strength at point P is

$$\cos\theta kI \sin(\omega t + a\cos\theta\cos\phi) \quad . \quad . \quad . \quad . \quad (8)$$

where ω is the angular frequency of the current I, and k is a constant of radiation.

Eqn. (8) can be expanded into a Bessel function series. Hence, field strength at P is

$$\begin{aligned} &kI\cos\theta\sin\omega t[J_0(a\cos\theta) - 2J_2(a\cos\theta)\cos 2\phi \\ &+ 2J_4(a\cos\theta)\cos 4\phi \ldots] \\ &+ kI\cos\theta\cos\omega t[2J_1(a\cos\theta)\cos\phi - 2J_3(a\cos\theta)\cos 3\phi \\ &+ 2J_5(a\cos\theta)\cos 5\phi)\ldots] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9) \end{aligned}$$

The above complex polar diagram of eqn. (9) could provide first- and general *n*th-order bearing signals if ϕ were allowed to vary according to the procedure described in Section 1.2. However, the selection of a first-order, and any one given *n*th-order, signal from a number of signals introduces practical difficulties, and it is preferable to eliminate the redundant signals by the following techniques.

(12.2.2) Polar Diagram of *n* Co-phased Dipoles.

Let each of the dipoles, Nos. 3, 5, 7 and 9, be fed with current I, and then added to element No. 1; each dipole is spaced by 72° from its immediate neighbour as shown.

At P the sum of the field strengths from the individual dipoles can be found from the addition of five equations of the form of eqn. (9) with ϕ replaced by ϕ, $\phi + 72°$, $\phi + 144°$, etc.

From the addition, it is shown that the total combined field strength is

$$\begin{aligned} &[5kI\cos\theta\sin\omega t J_0(a\cos\theta) \\ &+ 10kI\cos\theta\cos\omega t J_5(a\cos\theta)\cos 5\phi \\ &- 10kI\cos\theta\sin\omega t J_{10}(a\cos\theta)\cos 10\phi \ldots] \quad . \quad (10) \end{aligned}$$

It should be noted that all terms between the zero and fifth orders have disappeared.

The higher terms $J_{10}(a\cos\theta)$, etc., may be neglected if a is smaller than the value required to enable $J_5(a\cos\theta)$ to reach its maximum value. Throughout this Appendix such higher-order terms will be neglected, as a will be smaller than a_{max}.

In general, by using n elements, all being fed in phase, it can be shown similarly that the combined field strength at P is

$$nkI\cos\theta[J_0(a\cos\theta) + 2j^n J_n(a\cos\theta)\cos n\phi] \quad . \quad . \quad (11)$$

where the high-order terms of $J_{2n}(a\cos\theta)$, etc., have been neglected, and the time-dependent term $\sin\omega t$ has been omitted.

The above polar diagram consists of a zero-order plus an *n*th-order polar diagram. It is rarely convenient, for electrical and mechanical reasons, to produce both polar diagrams from the same array, and thus further diagram separation is often required.

(12.2.3) The Zero-Order Polar Diagram.

Let there be a very large increase in the number of dipoles around the array, and let each dipole be fed with a current I. Then the distant field strength at the point P can be written as

$$k_1 I\cos\theta J_0(a\cos\theta) \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where $k_1 = nk$.

The higher-order terms have now been completely removed since the number of aerials, n, around the circumference, $2\pi a$, can be made sufficiently large for the term $J_n(a\cos\theta)$ to be negligible.

Thus a zero-order, or omni-directional, polar diagram has been produced from one aerial system without any other order of diagram being present.

(12.2.4) The *n*th-Order Polar Diagram.

Let dipoles Nos. 2, 4, 6, 8 and 10, be added to dipoles Nos. 1, 3, 5, 7 and 9, as shown in Fig. 2, but let the new set be fed with the anti-phase current $-I$. The combined radiation from the two sets of dipoles can be found by adding two equations of the form of eqn. (10) with I and ϕ, and $-I$ and $(\phi + 36°)$, acting as the appropriate parameters.

Therefore the field strength at the distant point P is

$$= j20\,kI\cos\theta J_5(a\cos\theta)\cos 5\phi \quad . \quad . \quad . \quad (13)$$

Note that the zero-order term is no longer present.

In general, the *n*th-order polar diagram can be found from the expression:

Distant field strength at P

$$= j^n 4nkI\cos\theta J_n(a\cos\theta)\cos n\phi \quad . \quad . \quad . \quad (14)$$

The first-order polar diagram can be produced by using only the two aerials Nos. 1 and 6; i.e. $n = 1$.

(12.3) Polar Diagrams produced by Vertical Slots in Vertical Cylinders

The calculation of radiation from an infinitely long vertical lot in an infinitely long metal cylinder has already been considered.[13] However, an outline of the treatment will be given ere, since a number of intermediate steps in the analysis will be equired for the treatment of waveguide transducer polar liagrams.

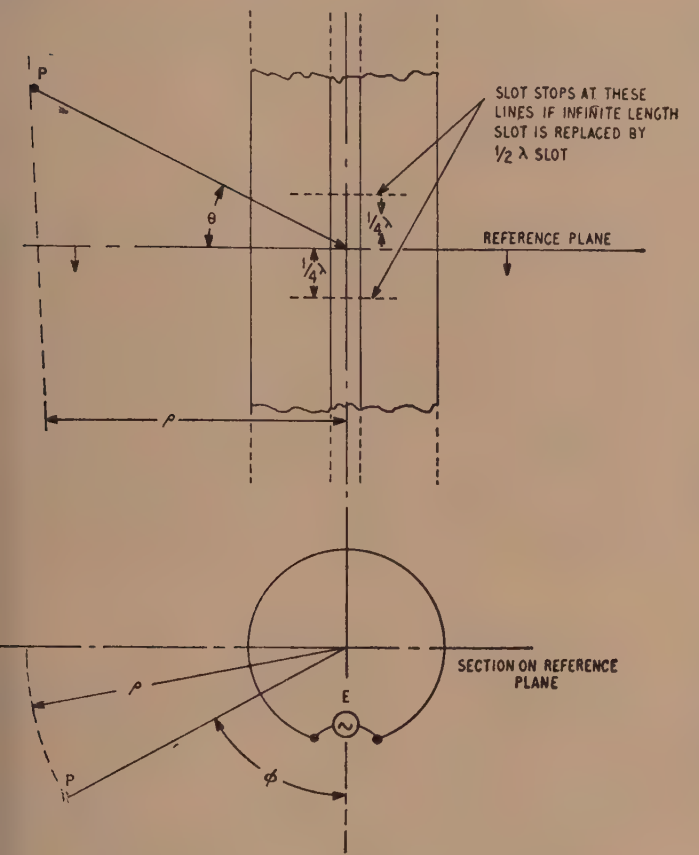


Fig. 13.—Slotted metal cylinder of infinite length.

Fig. 13 shows an infinitely long slot in an infinitely long ylinder of radius a. It has been shown[13] that the electric field E_ϕ, at the vertical line through the given point P, is given by

$$E_\phi = \sum_{p=-\infty}^{p=\infty} b_p \dot{H}_p^{(2)}(\rho)\, \varepsilon^{jp\phi} \quad . \quad . \quad . \quad . \quad (15)$$

where

ρ = Distance in radians between the axis of the cylinder and he vertical line through P.

ϕ = Angle between the plane containing the axis of the ylinder and the point P, and the plane through the axis bisecting he slot (see Fig. 13).

p = An integer.

$\dot{H}_p^{(2)}$ = First derivative of the pth-order Hankel function of he second kind.

b_p = A constant depending on the electric-field distribution round the circumference of the cylinder.

It has been shown[13] that, if the slot has an electric field E etween the two adjacent edges, b_p is given by

$$b_p = \frac{cE}{\dot{H}_p^{(2)}(a)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

here c is a constant depending on the slot width.

Again, it has been shown that, if ρ is large ($>10a$), it is permissible to rewrite eqn. (15) as

$$E_\phi = ckE \sum_{p=-\infty}^{p=\infty} \frac{\varepsilon^{jp\left(\phi+\frac{\pi}{2}\right)}}{\dot{H}_p^{(2)}(a)} \quad . \quad . \quad . \quad . \quad (17)$$

where k is a constant, and b_p has been substituted from eqn. (16).

The field strength at the distant point P from an energized $\frac{1}{2}\lambda$ slot in an infinitely long metal cylinder can be shown[13] to be expressed by eqn. (17), provided the radius factor a is replaced by $a \cos\theta$, where θ is the angle of elevation of the distant point P from a horizontal reference plane through the centre of the $\frac{1}{2}\lambda$ slot as is shown in Fig. 13, the angle being measured in the plane containing P and the vertical axis of the cylinder.

Therefore, for the $\frac{1}{2}\lambda$ slot,

$$E_\phi = ckE \sum_{p=-\infty}^{p=\infty} \frac{\varepsilon^{jp\left(\phi+\frac{\pi}{2}\right)}}{\dot{H}_p^{(2)}(a\cos\theta)} \quad . \quad . \quad . \quad (18)$$

It is now possible to find the polar diagrams produced by $\frac{1}{2}\lambda$ slots in metal cylinders when the slots are geometrically arranged in accordance with the descriptions of Section 12.2.

(12.3.1) The Zero-Order Polar Diagram

By using eqn. (18) and adopting the same approximations as given in Section 12.2.3, it is possible to find the distant field strength produced by a very large number of $\frac{1}{2}\lambda$ slots which are all fed in phase, i.e.

$$E_\phi = \frac{ck_1E}{\dot{H}_0^{(2)}(a\cos\theta)} \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where k_1 is a new constant (cf. Section 12.2.3).

(12.3.2) The nth Order Polar Diagram

By using eqn. (18) and adopting the technique of analysis used in Section 12.2.4, it is possible to find the distant field strength produced by $2n$ slots which are so fed that each has an electric-field excitation differing by 180° from the adjacent slot, i.e.

$$E_\phi = \frac{2nj^n ckE\cos n\phi}{\dot{H}_n^{(2)}(a\cos\theta)} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

(12.4) Polar Diagrams Produced by an Ideal Transducer

(12.4.1) General

That part of Fig. 5 which shows a section XX of the transducer can be used to represent a cross-section of an exciter-array cylinder which is infinitely long and has infinitely long slots and a transducer that has its waveguides replaced by radial conducting plates of infinite length. The arrangement will act as a perfect transducer array provided the following conditions are fulfilled:

(*a*) Only forward travelling waves exist between the exciter array and the outside circumference of the transducer (i.e. no reflections exist in the transducer inter-plate space, and in the space Δa between the exciter array and transducer).

(*b*) Each slot, defined by the gap between the two outer ends of the radial plates, can be regarded as an infinitely long radiating slot in a metal cylinder of radius a radians.

(*c*) An infinite number of samples of the exciter array field are made.

A transducer constructed according to the details given in the paper may not always satisfy these conditions. Considerable practical and theoretical work would be required before a complete theory could be developed to take into account the nonfulfilment of the conditions. However, with regard to (*c*), it is probable that the number of field samples would only have to

be equal to $2n$. The reason is that the sampling of the nth-order field variation must be similar to the pulse sampling of a sine wave where two samples per cycle are sufficient to define that wave. In practice, at least $7n$ field samples have been used, so it is expected that (*c*) has been fulfilled in the work described.

The performance of an ideal transducer may now be calculated. Assume that the exciter array of radius a has only one slot—the 'considered slot', as shown—and let this slot be fed with an electric field E. An electric field will exist on the outer circumference of the transducer and can be expressed in a general manner by eqn. (15), since the outside of the transducer can be regarded as a multi-slotted cylinder as stated in (*b*) above, i.e. at the outer circumference of the transducer,

$$E_\phi = \sum_{p=-\infty}^{p=\infty} b_{p\,trans}\dot{H}_p^{(2)}(a)\varepsilon^{jp\phi} \quad . \quad . \quad . \quad . \quad (21)$$

where $b_{p\,trans}$ is a constant similar to b_p of Section 10.3.

However, the field on the outer circumference is also equal to that on the inner circumference, as implied by (*a*) above. Thus the problem of finding the field on the outer circumference resolves into one of finding the field at a radius $a_s + \Delta a$ from a slotted cylinder of radius a_s.

Eqns. (15) and (16) may now be used to find this field at the radius $a_s + \Delta a$:

$$E_\phi = \sum_{p=-\infty}^{p=\infty} \frac{cE\varepsilon^{jp\phi}}{\dot{H}_p^{(2)}(a_s)}\dot{H}_p^{(2)}(a_s + \Delta a) \quad . \quad . \quad . \quad (22)$$

By equating eqns. (21) and (22),

$$b_{p\,trans} = \frac{\dot{H}_p^{(2)}(a_s + \Delta a)cE}{\dot{H}_p^{(2)}(a_s)\dot{H}_p^{(2)}(a)} \quad . \quad . \quad . \quad . \quad (23)$$

It is now possible to repeat the analysis given in Section 10.3, starting from the statement of the value of b_p and ending at the derivation of the zero- and nth-order polar diagrams for the $\frac{1}{2}\lambda$ slots in metal cylinders. This analysis is not repeated here, but the results are given below for the ideal transducer of $\frac{1}{2}\lambda$ height shown in Fig. 5:

For a zero-order polar diagram,

$$E_\phi = \frac{ck_1E}{\dot{H}_0^{(2)}(a\cos\theta)}\left[\frac{\dot{H}_0^{(2)}(a_s + \Delta a)}{\dot{H}_0^{(2)}(a_s)}\right] \quad . \quad . \quad . \quad (24)$$

For an nth-order polar diagram,

$$E_\phi = \frac{2nj^nckE}{\dot{H}_n^{(2)}(a\cos\theta)}\left[\frac{\dot{H}_n^{(2)}(a_s + \Delta a)}{\dot{H}_n^{(2)}(a_s)}\right]\cos n\phi \quad . \quad . \quad (25)$$

(12.4.2) The Transducer Factor

Eqns. (24) and (25) are similar to the slotted cylinder eqns. (19) and (20). The difference is in the additional factors such as

$$\frac{\dot{H}_0^{(2)}(a_s + \Delta a)}{\dot{H}_0^{(2)}(a_s)}$$

The additional factors are a measure of efficiency of the transducer in giving a small-radius body the radiation characteristics of a large-radius body. These are termed transducer factors, with notation t_0 and t_n, i.e.

$$t_0 = \frac{\dot{H}_0^{(2)}(a_s + \Delta a)}{\dot{H}_0^{(2)}(a_s)} \quad . \quad . \quad . \quad . \quad . \quad (26)$$

$$t_n = \frac{\dot{H}_n^{(2)}(a_s + \Delta a)}{\dot{H}_n^{(2)}(a_s)} \quad . \quad . \quad . \quad . \quad . \quad (27)$$

As would be expected, when $\Delta a = 0$, the factors become equal to unity, and full radius improvement has been given to the small rotating body.

Eqns. (26) and (27) can be used to express the radiation resistance of the exciter array, when in the transducer, compared with a slotted-cylinder array of outside radius a:

$$\text{Relative radiation resistance} = t_n^2 \quad . \quad . \quad . \quad (28)$$

since, if the slots of the two arrays are fed with an equal electric field, the relative radiation resistance depends on the square of the distant field.

(12.4.3) The Analysis of the Vertical Coverage of the First-Order Bearing Signal.

It should be noted that the transducer factor does not affect the shape of each polar diagram being transmitted. In other words, all the shapes of the diagrams can be calculated in terms of the radius a of the transducer without knowledge of the exact values of transducer factor, this being apparent from an inspection of eqns. (24) and (25). This concept is important when dealing with the transducer cavity of Fig. 10 (see Section 6), since it can then be seen that the loop and dipole exciter arrays of the transducer will be responsible for a distant field strength that is proportional to the amplitudes of the zero- and first-order polar diagrams of the slotted-cylinder arrays considered in Sections 12.3.1 and 12.3.2. The factors of proportionality will depend on the transducer factors of the loop, dipole and cavity arrangement. Thus the distant field strength produced by this transducer is

$$\left[\frac{A}{\dot{H}_0^{(2)}(a\cos\theta)} + \frac{B\cos\phi}{\dot{H}_1^{(2)}(a\cos\theta)}\right]\sin\omega t \quad . \quad . \quad (29)$$

where the constants A and B include the general radiation constants (proportional to the r.f. currents in the loop and dipole) as well as the transducer constants. The time-dependent term $\sin\omega t$ has been included.

Eqn. (29) contains the first- and zero-order r.f. signals and is representative of a bearing-modulated carrier wave as indicated in Section 1.2.1. The coefficient of amplitude modulation, $\cos\phi$, on the carrier is equal to the ratio B_1/A_1, where B_1 is the amplitude of the component of the first-order r.f. signal that is in r.f. phase or anti-phase with the zero-order r.f. signal, and $A_1 = |A/\dot{H}_0^{(2)}(a\cos\theta)|$ is the amplitude of the zero-order r.f. signal (carrier), i.e. the modulation coefficient is the real part of the ratio of first- to zero-order r.f. signal.

It is convenient if, in the ground plane of $\theta = 0$, the r.f. phase difference between the two signals is reduced to zero by the adjustment of the relative r.f. phase of the loop and dipole currents. This will mean that, for given amplitudes of the two r.f. signals, the modulation coefficient will be a maximum.

Mathematically, this will require that the ratio of the two signals has no imaginary component, i.e. the ratio

$$\frac{B}{\dot{H}_1^{(2)}(a)} \times \frac{\dot{H}_0^{(2)}(a)}{A} = m \text{ (a real constant)}$$

This implies that the ground-level modulation coefficient is m.

At a higher angle of elevation, θ, the ratio of the two signals becomes complex owing to an r.f. phase difference between the two signals, and the modulation coefficient at θ becomes

$$\mathscr{R}\frac{B\dot{H}_0^{(2)}(a\cos\theta)}{A\dot{H}_1^{(2)}(a\cos\theta)} = \mathscr{R}m\frac{\dot{H}_1^{(2)}(a)\dot{H}_0^{(2)}(a\cos\theta)}{\dot{H}_0^{(2)}(a)\dot{H}_1^{(2)}(a\cos\theta)}$$

Therefore the modulation coefficient at θ

$$= m\left|\frac{\dot{H}_1^{(2)}(a)\dot{H}_0^{(2)}(a\cos\theta)}{\dot{H}_0^{(2)}(a)\dot{H}_1^{(2)}(a\cos\theta)}\right|\cos\chi \quad . \quad . \quad . \quad (30)$$

where

$$\chi = \arg\frac{\dot{H}_0^{(2)}(a)}{\dot{H}_1^{(2)}(a)} - \arg\frac{\dot{H}_0^{(2)}(a\cos\theta)}{\dot{H}_1^{(2)}(a\cos\theta)} \quad . \quad . \quad . \quad (31)$$

χ is thus the r.f. phase difference between the two signals at the elevation θ.

Eqns. (30) and (31) enable the fall in modulation coefficient to be calculated relative to a maximum value of that at ground level ($\theta = 0$), and can thus be used to calculate the vertical-coverage angle as described in Section 7.

It should be noted that the Hankel functions in eqn. (29) have only one zero, which occurs when the arguments are zero ($\theta = 90°$); this means that the polar diagram expressed by eqn. (29) is free from multiple amplitude nulls in the vertical plane. This desirable freedom does not exist for the vertical dipole arrays of Section 12.2, as can be seen from the dependence of each diagram on a single Bessel function which must have multiple zeros if the argument (proportional to the array radius) is made large.

621.372.543:621.375.4

The Institution of Electrical Engineers
Paper No. 2787 E
Mar. 1959
©

TRANSISTOR ACTIVE FILTERS USING TWIN-T REJECTION NETWORKS

By A. E. BACHMANN, Dipl.El.Ing.

(*The paper was first received* 11*th August, and in revised form* 1*st October*, 1958.)

SUMMARY

The introduction of a rejection-type network, such as a twin-T or a bridged-T, into the feedback loop of an amplifier leads to bandpass characteristics similar to those of single tuned circuits. It is possible to achieve highly selective circuits at very low frequencies, e.g. in the audio range, where high-Q-factor coils are usually physically large and expensive. Some basic circuits using a symmetrical twin-T network with R and C only in the feedback loop of a transistor amplifier are investigated. Results of measurements are given. Q-factors higher than 50 are obtained at a frequency of 600 c/s.

LIST OF SYMBOLS

A_v = Open loop voltage gain.
A_0 = Voltage gain of transistor stage, when terminated with load (y_L) in parallel with twin-T (y''_{22}).
A_T = Voltage gain of twin-T, when terminated with the transistor (y'_{22}) in parallel with the load (y_L).
a_{ik}, h_{ik}, y_{ik} = a-, h- and y-parameters of four-pole network respectively.
Subscript b: Common-base connection.
Subscript e: Common-emitter connection.
Subscript c: Common-collector connection.
r = Function of frequency, having a sharp resonance.
p = $j\omega$, imaginary frequency.
F = Transfer function of a four-terminal network (output voltage divided by generator voltage).
Q = Q-factor of a coil, single tuned circuit, or selective circuit using RC rejection-type networks in the feedback loop.
R = Resistances in twin-T network.
R_g = Generator resistance.
R_L = $1/y_L$: Load resistance.
β = v_2/v_1: Transmission of twin-T network.
Δ_y, Δ_h = Determinants of the y-matrix and h-matrix.
ζ = Parameter of function r which is characteristic for the resonance.
τ = Time-constant.
ω = Angular frequency.

(1) INTRODUCTION

Low-frequency narrow-band filters require large coils having only small losses. Active elements, such as transistors and vacuum tubes, can, however, be used to compensate for all or part of the losses, and thus make it possible to use coils of a smaller size. This also opens up the possibility of using RC networks exclusively, for example in the feedback loop of an amplifier. A number of authors have written on the subject.[1–11] The paper presents an analysis of an active low-frequency filter using a symmetrical twin-T network with R and C only in the feedback loop of a transistor amplifier. The main objective is to design an amplifier which will achieve Q-factors up to 100. Several possible circuits will be discussed, and the Q-factors which can be expected from them will be given.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Bachmann is at the P.T.T. Research Laboratory, Switzerland.

(2) GENERAL PROBLEM

A typical feedback circuit is shown in Fig. 1. The amplifier is represented by its y-matrix, y', and in parallel with it is the feedback network, which is also represented by its y-parameters,

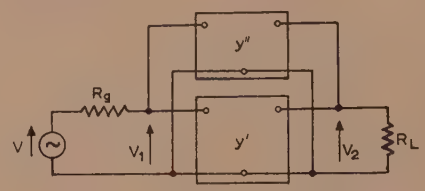


Fig. 1.—General feedback circuit.

y''. Since the two 4-terminal networks are in parallel, the overall y-matrix is the sum of the two single y-matrices. It can be shown that the transfer function of interest is given by

$$F = \frac{-V_2}{V} = \frac{y_{21}}{y_{22} + y_L + (\Delta_y + y_{11}y_L)R_g} \quad . \quad . \quad (1)$$

Eqn. (1) is developed in Appendix 9.1 for typical values of the different y-parameters, and the Q-factor of the circuit in question is calculated to be

$$Q \simeq \frac{y'_{21}}{2(\sqrt{2} + Ry'_{11} + R/R_g)(y'_{22} + y_L + \sqrt{2}/R)} \quad . \quad (2)$$

The optimum value of R for the twin-T network is given by

$$R_{opt} \simeq \sqrt{\left(\frac{2}{y_L y'_{11}}\right)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

From eqn. (2) it can be seen that, in order to achieve high Q-factors, the following steps must be taken:

(*a*) Make y'_{21} large.
(*b*) Choose $R \simeq R_{opt}$ as given by eqn. (3).
(*c*) Make R_g large and y'_{11}, y'_{22}, y_L small.

(3) VARIOUS CIRCUIT POSSIBILITIES

Four typical circuits are shown in Fig. 2. The y'-parameters are given in Table 1 and compared with those of a vacuum tube; also the Q-factor is calculated and compared with measured values. The y'-parameters for the common-emitter common-collector (C.E.C.C.) compound circuit of Fig. 2(*c*) are calculated in Appendix 9.3.

It will be seen that the circuits in Figs. 2(*c*) and (*d*) are superior to the others, since they lead to higher Q-factors. The main reason for this is the higher value of y'_{21}. With a proper choice of transistors and d.c. bias, the circuit in Fig. 2(*c*), which uses only two transistors, is capable of giving Q-factors of over 50. With a three-transistor circuit, such as the one shown in Fig. 2(*d*), it should be possible to get Q-factors of 100 or more.

(4) CIRCUITS WITH HIGH Q-FACTORS

While the circuits of Figs. 2(*c*) and (*d*) both have a Q-factor greater than 50, the two-transistor circuit [Fig. 2(*c*)] has the

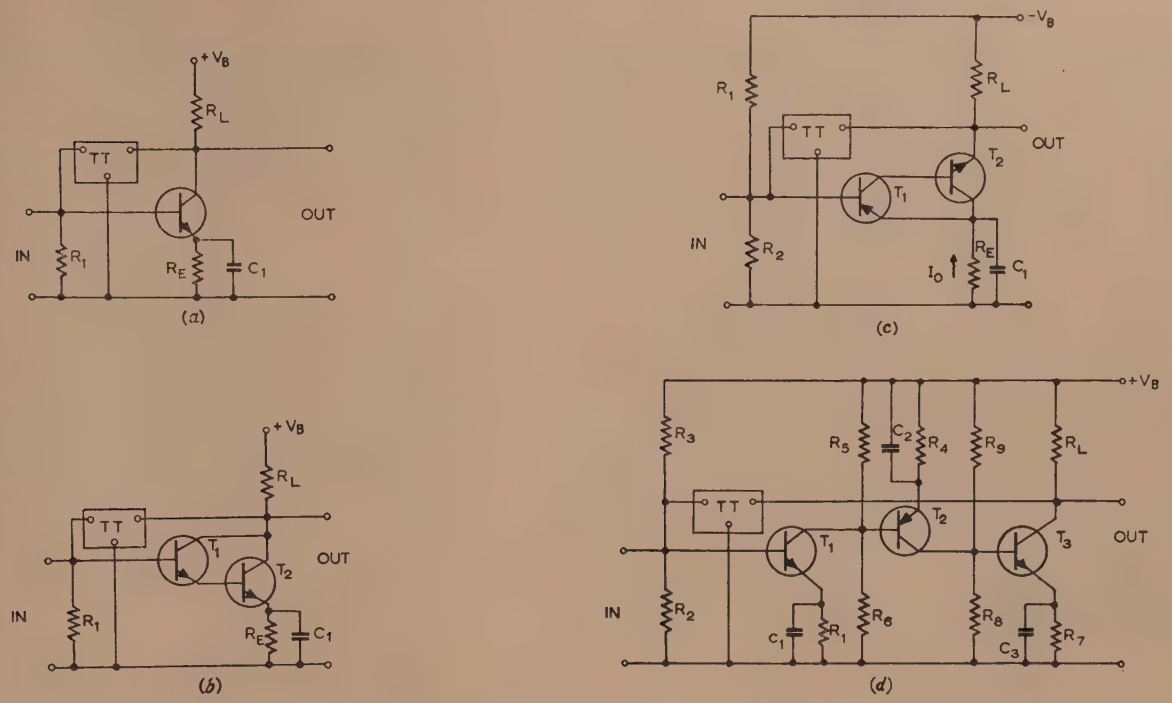


Fig. 2.—Typical circuits.

(*a*) Common-emitter stage.
(*b*) Darlington compound connection.
(*c*) C.E.C.C. compound connection.
(*d*) Three common-emitter stages.

Table 1

y′-PARAMETERS IN MHOS AND Q-FACTORS OF THE CIRCUITS SHOWN IN FIG. 2

	Vacuum tube	Circuit in Fig.:			
		2(*a*)	2(*b*)	2(*c*)	2(*d*)
y'_{11}	10^{-6}	5×10^{-4}	6×10^{-6}	$7{\cdot}3 \times 10^{-5}$	5×10^{-4}
y'_{12}	0	$-8{\cdot}5 \times 10^{-8}$	-8×10^{-9}	$-1{\cdot}8 \times 10^{-7}$	<(2*a*)
y'_{21}	$5 \times 10^{-3} A_v$	2×10^{-2}	10^{-2}	$7{\cdot}5 \times 10^{-2}$	>(2*a*)
y'_{22}	10^{-4}	$6{\cdot}6 \times 10^{-6}$	5×10^{-6}	$5{\cdot}2 \times 10^{-5}$	$6{\cdot}6 \times 10^{-6}$
Calculated Q-factor		4·1	15	50	100
Measured Q-factor		4	10	55	60

great advantage that it is stable as regards oscillation, owing to the phase shift in the amplifier, whereas the three-transistor circuit [Fig. 2(*d*)] tends to oscillate. Both circuits have the following disadvantages (see also Fig. 1):

(*a*) Since the input impedance of the amplifier should not load the twin-T network too heavily, the source resistance R_g must be large.
(*b*) In order to get most of the signal available from the amplifier into the twin-T network for high feedback, the load resistance R_L should not be too small.

Because of these two disadvantages another circuit was developed (Fig. 3). It makes use of the amplification of two ordinary common-emitter stages, and the twin-T network is terminated by a common-collector stage. The load on this stage is essentially the input impedance of the first transistor, T_1, in common-base connection and consequently it is very low. In order to increase the input impedance of the third stage, the resistance R_{E3} has been added to the load.

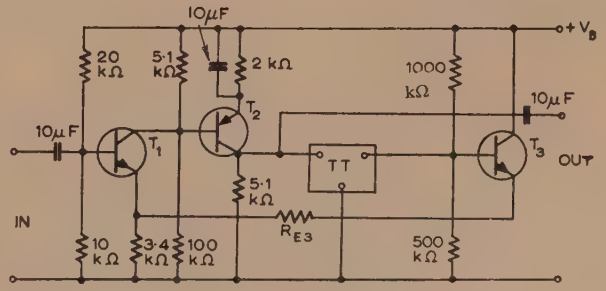


Fig. 3.—High Q-factor circuit.

(5) MEASUREMENTS

The performance of the circuit in Fig. 2(*c*) was measured at 600 c/s. Table 2 gives the measured values of h'_{21}, y'_{11}, and $y'_{21} = h'_{21}y'_{11}$ for a few transistors.

Table 2

h'_{21}, y'_{11}, AND y'_{21} OF THE C.E.C.C. CIRCUIT, FOR $I_{E1} + I_{C2} = 2$mA

T_2	h_{21e2}	T_1	h_{21e1}	h'_{21}	y'_{11}	y'_{21}
					mhos $\times 10^{-5}$	mhos $\times 10^{-2}$
2N78	35	2N43	50	1030	7·0	7·2
2N78	35	2N137	70	2000	2·9	5·8
2N78	35	2N241A	90	1340	5·2	7·0
2N78	15	2N43	50	664	12·1	8·0
2N78	50	2N43	50	1460	4·5	6·6

The total current through the common-emitter resistor R_E was kept constant at 2 mA.

It should be noted that the collector current of T_1 is the base current of T_2 and is thus very small. On the other hand, it is desirable to operate T_1 at as high a current as possible, so that $y'_{21e} = 1/h'_{11b}$ becomes large. One way to increase the current in T_1 is to choose a transistor for T_2 which has a low current gain; an alternative is to connect a shunt resistor from the collector of T_1 to $-V_B$. Table 1 substantiates this, since the large value of y'_{21} appears when a transistor which has low value of h_{21e2} is used for T_2.

In order to obtain a Q-factor of approximately 50 the following values had to be used:

$R_L = 20\,\text{k}\Omega$ $\quad R_E = 5\,\text{k}\Omega$
$R_1 = \infty$ $\quad R_2 = 80\,\text{k}\Omega$
$R_g = 10\,\text{k}\Omega$ $\quad T_2 = 2\text{N}78$
$T_1 = 2\text{N}43$ $\quad I_0 = 3\,\text{mA}$
$V_B = -75\,\text{V}$

Since the requirements on the battery power are very severe, the circuit does not look very promising for portable applications.

The circuit of Fig. 3 was measured at 600 c/s with the following alterations: $R_{E3} = 0$, but T_3 was actually two transistors in a Darlington compound connection. There was no *RC* phase-shift network between the twin-T network and T_3, but there were large shunt capacitances to earth. For an open-loop voltage gain of $A_v = 220$ between the base of T_3 and the collector of T_2, the measured Q-factor was 54, which agrees with the formula[12] $Q \simeq A_v/4 = 55$.

(6) CONCLUSIONS

The investigation and measurements show that, in order to achieve a Q-factor greater than 50, a three-transistor amplifier is needed. For values of Q less than 50, a two-transistor amplifier may be sufficient if the proper transistor types and a high enough battery power are available. In order to achieve highly-selective circuits, the symmetrical twin-T's to be used must have a rejection of approximately 50–60 dB at the centre frequency. This means that stable components have to be used, which are expensive; furthermore, miniaturization is not very easy to achieve.

(7) ACKNOWLEDGMENT

I wish to thank the Electronics Laboratory of the General Electric Co., Syracuse, New York, where I had the opportunity to carry out this work.

(8) REFERENCES

(1) COWLES, R. G.: 'The Parallel-T R–C Network', *Proceedings of the Institute of Radio Engineers*, 1952, **40**, p. 1712.
(2) MCGAUGHAN, H. S.: 'Variation of an R–C Parallel-T Network', *Tele-Tech*, 1947, **6**, p. 48.
(3) HARRIS, G. R.: 'Bridged Reactance Resistance Networks', *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 883.
(4) HASTINGS, A. E.: 'Analysis of a Resistance Capacitance Parallel T Network and Application', *ibid.*, 1946, **34**, p. 126.
(5) OONO, J.: 'Design of Parallel T Resistance Capacitance Networks', *ibid.*, 1955, **43**, p. 617.
(6) SMITH, D. H.: 'The Characteristics of Parallel-T RC Networks', *Electrical Engineering*, 1957, **2**, p. 7.
(7) STANTON, L.: 'Theory and Application of Parallel T Resistance Capacitance Frequency Selective Networks', *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 447.
(8) WARD, J. J., and LANDSHOFF, P. V.: 'Parallel-T RC Selective Amplifiers', *Electronic and Radio Engineer*, 1958, **4**, p. 120.
(9) WOLF, A.: 'Note on Parallel T Resistance Capacitance Network', *Proceedings of the Institute of Radio Engineers*, 1946, **34**, p. 659.
(10) WHITE, C. F., and MORGAN, K. A.: 'The Dual-Input Parallel-T Network', Naval Research Laboratory Report, NRL 4011, July, 1952.
(11) WELLMAN, B.: 'Convenient Twin-T Circuits', *Electrical Design News*, October, 1956, p. 46.
(12) VALLEY, G. E., and WALLMAN, H.: 'Vacuum Tube Amplifiers', Radiation Laboratory Series, Vol. 18 (McGraw-Hill, 1948).

(9) APPENDICES

(9.1) Q-factor of the General Feedback Circuit

Referring to Fig. 1 and eqn. (1), the transfer function F of interest is given by

$$F = \frac{-V_2}{V_\cdot} = \frac{y'_{21} + y''_{21}}{y'_{22} + y''_{22} + y_L + R_g y_L(y'_{11} + y''_{11}) + R_g[(y'_{11} + y''_{11})(y'_{22} + y''_{22}) - (y'_{12} + y''_{12})(y'_{21} + y''_{21})]} \quad (4)$$

Since $y'_{21} \gg y''_{21}$, eqn. (4) may be written

$$F \simeq \frac{y'_{21}}{y'_{22} + y''_{22} + y_L + R_g y_L(y'_{11} + y''_{11}) + R_g[(y'_{11} + y''_{11})(y'_{22} + y''_{22}) - (y'_{12} + y''_{12})y'_{21}]} \quad . \quad . \quad (5)$$

Let A_0 be the voltage gain of the transistor when terminated by the load y_L and the short-circuited twin-T network y''_{22}. Because $y''_{11} = y''_{22}$, A_0 is given by

$$A_0 = \frac{-y'_{21}}{y'_{22} + (y_L + y''_{22})} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Substituting eqn. (6) in eqn. (5), F becomes

$$F \simeq \frac{A_0}{1 + R_g(y'_{11} + y''_{11}) - \dfrac{R_g y'_{21}(y'_{12} + y''_{12})}{y'_{22} + y_L + y''_{22}}} \quad . \quad . \quad (7)$$

Let A_T denote the voltage gain of the twin-T when it is terminated by the transistor y'_{22} and the load y_L. Since $y''_{12} = y''_{21}$,

$$A_T = \frac{-y''_{12}}{y''_{22} + y_L + y'_{22}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Substituting eqn. (8) in eqn. (7),

$$F \simeq \frac{A_0}{1 + R_g(y'_{11} + y''_{11}) - \dfrac{y'_{12}}{y'_{22} + y_L + y''_{22}} + A_T R_g y'_{21}} \quad . \quad (9)$$

With the aid of Table 1, it can be seen that in all cases

$$\frac{y'_{12}}{y'_{22} + y_L + y''_{22}} \ll 1$$

so that F is finally given by

$$F \simeq \frac{A_0/(R_g y'_{21})}{(y'_{11} + y''_{11} + 1/R_g)/y'_{21} + A_T} \quad . \quad . \quad . \quad (10)$$

In order to simplify the following calculations it will be assumed that, in the neighbourhood of the centre frequency f_0, all the terms in eqn. (10) are constant with the exception of y''_{12}. This leads to the following formula for the transfer function F:

$$F \simeq \frac{a}{b - y''_{12}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where a and b are given by

$$a = 1/R_g; \; b = 1/y'_{21}(y'_{11} + y''_{11} + 1/R_g)(y''_{22} + y_L + y'_{22}) \quad . \quad (12)$$

If the value of y''_{12} from eqn. (22) (see Appendix 9.2) is substituted in eqn. (11), F can be expressed as follows:

$$F \simeq \frac{\dfrac{2aR}{1 + 2bR}(1 + p\tau_0)}{1 + \dfrac{2bR}{1 + 2bR} p\tau_0 + \dfrac{p^2\tau_0^2}{1 + 2bR}} \quad . \quad . \quad . \quad (13)$$

Compare the function in the denominator of eqn. (13) with the function r in eqn. (14):

$$r = 1 + 2\zeta p\tau + p^2\tau^2 \quad . \quad . \quad . \quad . \quad . \quad (14)$$

For the case $\zeta \ll 1$, the Q-factor of r is defined by

$$Q = \frac{1}{2\zeta} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

If the Q-factor of the function in eqn. (14) is 100 ($Q \gg 1$), then ζ must be much less than 1, or $2bR \ll 1$, which means that

$$Q \simeq \frac{1}{2bR} = \frac{y'_{21}}{2R(y'_{11} + y''_{11} + 1/R_g)(y''_{22} + y_L + y'_{22})} \quad . \quad (16)$$

or, with $y''_{11} \simeq \dfrac{\sqrt{2}}{R}$ (see Fig. 6),

$$Q \simeq \frac{y'_{21}}{2(\sqrt{2} + Ry'_{11} + R/R_g)(y'_{22} + y_L + \sqrt{2}/R)} \quad . \quad (17)$$

This final formula for the approximate value of the Q-factor shows how to achieve a large one. It also makes possible the calculation of the optimum R of the twin-T. From eqn. (17), R_{opt} is found to be

$$R_{opt} = \sqrt{\left(\frac{2}{y_L y'_{11}}\right)}$$

The Q-factors as calculated for the circuits in Fig. 2 are given in Table 1. Compared with the measurements, they show the validity of eqn. (17) and the approximations which were made.

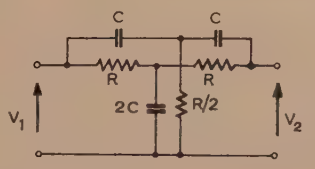


Fig. 4.—Symmetrical twin-T network.

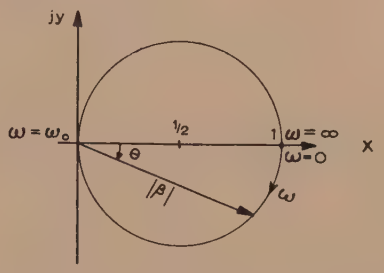


Fig. 5.—Transmission β of twin-T network.

(9.2) Symmetrical Twin-T (with R and C only)

The twin-T network of Fig. 4 is described in many articles.[1–11] It has, together with the bridged-T and the Wien bridge circuits, a transmission $\beta = v_2/v_1$ of the form given in eqn. (19) and shown in Fig. 5:

$$\beta = \frac{1}{1 - j\dfrac{4\omega/\omega_0}{(\omega/\omega_0)^2 - 1}} \quad . \quad . \quad . \quad . \quad (19)$$

The frequency of maximum rejection ($|\beta| = 0$) is given by

$$\omega_0 = \frac{1}{RC} = \frac{1}{\tau_0} \quad . \quad . \quad . \quad . \quad . \quad (20)$$

The y'' parameters can be calculated from Fig. 4. They are

$$y''_{11} = y''_{22} = \frac{1 + 4p\tau_0 + p^2\tau_0^2}{2R(1 + p\tau_0)} \quad . \quad . \quad . \quad (21)$$

$$y''_{12} = y''_{21} = \frac{1 + p^2\tau_0^2}{-2R(1 + p\tau_0)} \quad . \quad . \quad . \quad (22)$$

In Fig. 6 the absolute magnitudes of the y'' parameters are plotted against frequency. It will be seen that y''_{11} and y''_{22} are slowly varying functions of the frequency, whereas y''_{12} and y''_{21} change very rapidly in the neighbourhood of the centre frequency f_0.

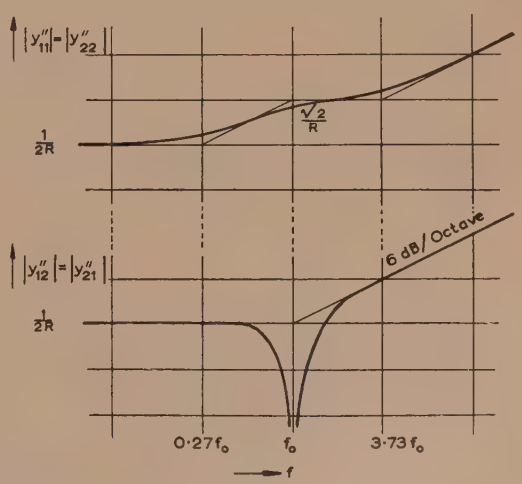


Fig. 6.—y-parameters of symmetrical twin-T network.

(9.3) y'-Parameters of the C.E.C.C. Circuit

The C.E.C.C. compound circuit in Fig. 2(*c*) consists of a common-emitter stage which is directly coupled to a common collector stage. The a.c. circuit is shown in Fig. 7. The two stages are in cascade, so that the *a*-matrix of the compound circuit is given by

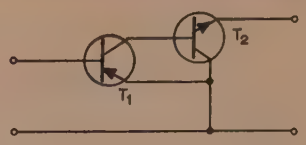


Fig. 7.—C.E.C.C. compound connection (a.c. circuit).

$$[a] = [a'] \cdot [a''] = \begin{bmatrix} \frac{\Delta'_h}{h'_{21}} & \frac{h'_{11}}{h'_{21}} \\ \frac{h'_{22}}{h'_{21}} & \frac{1}{h'_{21}} \end{bmatrix} \cdot \begin{bmatrix} \frac{\Delta''_h}{h''_{21}} & \frac{h''_{11}}{h''_{21}} \\ \frac{h''_{22}}{h''_{21}} & \frac{1}{h''_{21}} \end{bmatrix} = \frac{1}{h'_{21}h''_{21}} \begin{bmatrix} \Delta'_h\Delta''_h + h'_{11}h''_{22} & \Delta'_h h''_{11} + h'_{11} \\ h'_{22}\Delta''_h + h''_{22} & h'_{22}h''_{11} + 1 \end{bmatrix} \quad . \quad . \quad . \quad . \quad (23)$$

Δ'_h and Δ''_h being the determinants of the respective *h*-matrices. Transforming into *y*-parameters:

$$[y] = \frac{1}{a_{12}} \begin{bmatrix} a_{22} & -\Delta_a \\ -1 & a_{11} \end{bmatrix} = \frac{1}{\Delta'_h h''_{11} + h'_{11}} \begin{bmatrix} h'_{22}h''_{11} + 1 & \frac{-(\Delta'_h\Delta''_h + h'_{11}h''_{22})(h'_{22}h''_{11} + 1) + (\Delta'_h h''_{11} + h'_{11})(h'_{22}\Delta''_h + h''_{22})}{h'_{21}h''_{21}} \\ h'_{21}h''_{21} & \Delta'_h\Delta''_h + h'_{11}h''_{22} \end{bmatrix} \quad . \quad (24)$$

Note that in eqns. (23) and (24) the following notation is used:

$'$ common-emitter values of T_1
$''$ common-collector values of T_2

The common-collector values of T_2 are now converted into common-emitter values, with the help of eqns. (25):

$$\left.\begin{aligned} h''_{11c} &= h''_{11e} \\ h''_{12c} &= 1 - h''_{12e} \\ h''_{21c} &= -(1 + h''_{21e}) \\ h''_{22c} &= h''_{22e} \\ \Delta^c_h &= \Delta^e_h + 1 + h''_{21e} - h''_{12e} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

Substituting the values of eqn. (25) into eqn. (24),

$$\left.\begin{aligned} y_{11} &= \frac{h'_{22}h''_{11} + 1}{\Delta'_h h''_{11} + h'_{11}} \\ y_{12} &= \frac{[\Delta'_h(\Delta''_h + 1 + h''_{21} - h''_{12}) + h'_{11}h''_{22}](h'_{22}h''_{11} + 1) - (\Delta'_h h''_{11} + h'_{11})[h'_{22}(\Delta''_h + 1 + h''_{21} - h''_{12}) + h''_{22}]}{h'_{21}(1 + h''_{21})(\Delta'_h h''_{11} + h'_{11})} \\ y_{21} &= -\frac{h'_{21}(1 + h''_{21})}{\Delta'_h h''_{11} + h'_{11}} \\ y_{22} &= \frac{\Delta'_h(\Delta''_h + 1 + h''_{21} - h''_{12}) + h'_{11}h''_{22}}{\Delta'_h h''_{11} + h'_{11}} \end{aligned}\right\} \quad . \quad . \quad (26)$$

where all the values are expressed in terms of common emitter parameters.

The following approximations can be made:

$$\left.\begin{aligned} &h'_{22}h''_{11} \ll 1 \\ &\Delta'_h h''_{11} \ll h'_{11} \\ &\Delta''_h + 1 - h''_{12} \ll h''_{21} \\ &h''_{21} \gg 1 \\ &h'_{11}h''_{22} \ll \Delta'_h h''_{21} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

With these approximations, the *y*-parameters are given by

$$\left.\begin{aligned} y_{11} &\simeq 1/h'_{11} \\ y_{12} &\simeq -h'_{12}/h'_{11} \\ y_{21} &\simeq -h'_{21}h''_{21}/h'_{11} \\ y_{22} &\simeq \Delta'_h h''_{21}/h'_{11} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

Typical values for these *y*-parameters are given in Tables 1 and 2.

621.372.83

The Institution of Electrical Engineers
Paper No. 2797 E
Mar. 1959
©

THE TRANSFORMATION OF ADMITTANCE THROUGH A MATCHING SECTION AND LOSSLESS WAVEGUIDE JUNCTION

By J. R. G. TWISLETON, B.Sc., Associate Member.

(*The paper was first received 22nd April, and in revised form 9th October*, 1958.)

SUMMARY

The paper describes a method of setting up a matching section, comprising a variable reactor and phase shifter, in cascade with a lossless waveguide junction to obtain a particular admittance transformation. The terminal planes are defined by the position of voltage zeros when the junction is terminated by a short-circuit. The matching section and junction are regarded as two junctions in cascade and, using conventional theory, the relation between the admittance transformation parameters and input-voltage reflection coefficients is derived. Some experimental results obtained using an S-band waveguide system are quoted to demonstrate the application of the theory.

LIST OF PRINCIPAL SYMBOLS

r', r'' = Voltage reflection coefficients at junction input and output terminal planes.
y', y'' = Normalized admittance at junction input and output terminal planes.
α, R, β = Parameters of bilinear transformation relating voltage reflection coefficients.
$\theta/2$ = Electrical length of waveguide between ideal transformers in equivalent circuit.
A, B = Parameters of admittance transformation.
γ = Function of R where $R = \tan(\gamma/2)$.
S = Voltage standing-wave ratio, defined > 1.

(1) INTRODUCTION

Experimental waveguide junctions, i.e. transitions between guides of different sizes or shapes, seldom have precisely the required impedance transformation properties. Therefore it is often convenient in practice to modify the transmission properties of a junction by the use of a matching section[1,2] in cascade with the junction. The matching section may be used to obtain zero reflection, or to obtain a particular impedance or admittance transformation.

Matched junctions are usually required in transmission systems, but certain components, for example crystal holders and waveguide output circuits for microwave valves, are designed to transform impedances in a definite ratio. In this case it is necessary, of course, to define terminal planes for which the transformation is appropriate, since different transformations are obtained between different pairs of terminal planes.

The theoretical relation between the transformed impedances is a bilinear function in which three complex constants specify the function properties.[3] These constants, or the scattering coefficients of the junction,[3] can be found by calculation or by graphical methods, from data obtained using a short-circuit termination.[4,5,6] The derivation of these parameters is a necessary preliminary to the calculation of corresponding parameters for the matching section.

If, however, the junction is intended to transform the characteristic impedance of the waveguide at one side of the junction to a particular impedance value at the other side, or vice versa, the position of only one terminal plane is important. It is then possible to choose one of a pair of terminal planes, defined for a lossless junction by the positions of zero voltage when the junction is terminated by a short-circuit, to coincide with the specified terminal plane. The terminal and input admittances are linearly related and only two constants are required to specify the transformation[7] at any particular frequency. This choice of terminal planes simplifies the calculation involved in setting up the matching section.

Here we consider the use of a matching section consisting of a variable reactor and phase shifter;* the latter is inserted between the reactor and junction.

The only important parameters of this matching section are the reflection produced by the reactor when terminated in a matched load, and the phase shift introduced between the reactor and the junction. These can be found by independent measurement. We assume that the phase shifter produces no reflection, and that any phase shift introduced by the reactor is independent of the magnitude of reflection. The first of these idealizations can almost be achieved in practice, whilst phase shift by the reactor can be allowed for.

The paper describes a method of setting up this type of matching section when only a short-circuit termination is available. This case is of practical importance, since it is often easier to provide a short-circuit rather than a matched termination when the waveguide is a special type, e.g. a ridge guide.

(2) THE TRANSFORMATION OF VOLTAGE REFLECTION COEFFICIENT AND ADMITTANCE

We first consider the transformation of the voltage reflection coefficient through one junction, and then through two junctions in cascade, to obtain relations between the junction parameters in these transformations and those in the corresponding admittance transformations.

(2.1) Transformation of Voltage Reflection Coefficient

It is well known that the bilinear transformation relating voltage reflection coefficients at the output and input terminal planes of a lossless junction has the form[8]

$$r' = \varepsilon^{j\alpha}\frac{R + \varepsilon^{j\beta}r''}{1 + R\varepsilon^{j\beta}r''} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

since it must transform unit reflection coefficient, $\varepsilon^{j\phi}$ say, corresponding to a reactive termination, without change of magnitude. In eqn. (1), R is the magnitude of the input-voltage reflection coefficient of the junction with a matched termination, whilst α and β are phase constants which depend upon the positions of the input and output terminal planes. Fig. 1(*a*) shows an equivalent circuit for the junction, which is evident from eqn. (1). Fig. 1(*c*) shows a similar circuit for two junctions in cascade in

* A reactor which can be moved relative to the junction is equally suitable.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Twisleton is at the Research Laboratory of the British Thomson-Houston Co., Ltd.

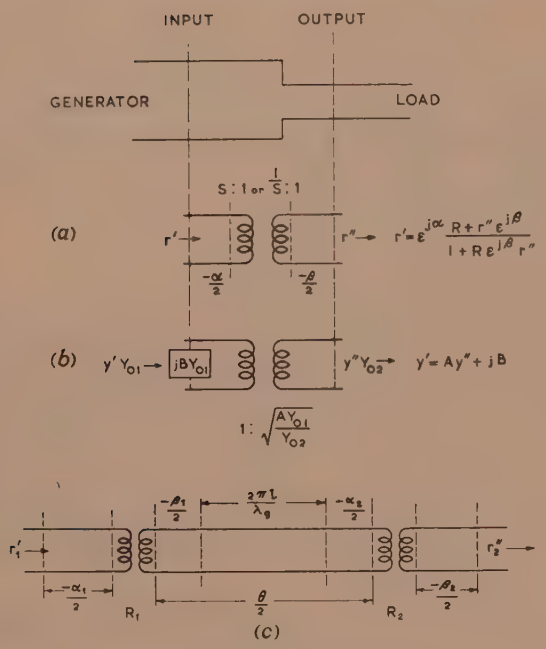


Fig. 1.—Equivalent circuits.
(*a*), (*b*) For lossless waveguide junction.
(*c*) For matching section and lossless waveguide junction.

which a transmission line of length $\theta/2$ separates the two ideal transformers. It is possible to reduce this circuit to that of Fig. 1(*a*), as shown in Appendix 8.2, for which the parameters α, R and β are

$$R = \sqrt{\left(\frac{R_1^2 + R_2^2 + 2R_1R_2\cos\theta}{1 + R_1^2R_2^2 + 2R_1R_2\cos\theta}\right)} \quad . \quad . \quad . \quad (2)$$

$$\alpha = \alpha_1 + \arctan\frac{R_2(R_1^2 - 1)\sin\theta}{R_1(1 + R_2^2) + R_2(1 + R_1^2)\cos\theta} \quad . \quad (3)$$

$$\beta = \beta_2 + \arctan\frac{R_1(R_2^2 - 1)\sin\theta}{R_2(1 + R_1^2) + R_1(1 + R_2^2)\cos\theta} \quad . \quad (4)$$

where the subscripts relate to junctions 1 and 2. So far, the position of the two terminal planes has been arbitrary, but we now define one of them by the position of a voltage zero with a short-circuit termination at the other.

(2.2) Transformation of Admittance

With the above definition of terminal planes, the normalized input and terminal admittances y' and y'' are linearly related. We can put

$$y' = Ay'' + jB \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where A and B are real numbers, since a pure susceptance must be transformed into another.* These constants depend upon the position of the pair of terminal planes; they can be found from a shift of a voltage zero when a short-circuit is displaced $\lambda_g/8$ and $\lambda_g/4$ in turn from the output terminal plane. If l_1 and l_2 are the corresponding shifts in the position of the input voltage zero, then $B - A = \cot(2\pi l_1/\lambda_g)$, and $B = \cot(2\pi l_2/\lambda_g)$. Alternatively, the values of $B - A$ and B can be found directly from the Smith chart, measuring distances l_1/λ_g and l_2/λ_g in an anticlockwise direction from the point (j, ∞) on the unit circle.

* Eqn. (5) suggests an equivalent circuit for the junction consisting of an ideal transformer with a shunt input susceptance as shown in Fig. 1(*b*). This circuit is a simplification of a more general one.[9]

Using the equations $r' = \dfrac{1 - y'}{1 + y'}$, and $y'' = \dfrac{1 - r''}{1 + r''}$ it is easy to derive the expression relating r' and r'' similar to eqn. (1) in terms of A and B. We find, with the aid of eqns. (30) of Appendix 8.1, that

$$R^2 = \frac{(1 - A)^2 + B^2}{(1 + A)^2 + B^2} = \frac{1 - \dfrac{2A}{1 + A^2 + B^2}}{1 + \dfrac{2A}{1 + A^2 + B^2}} \quad . \quad (6)$$

$$\alpha = \arctan\frac{-2B}{1 - A^2 - B^2} \quad . \quad . \quad . \quad . \quad (7)$$

and

$$\beta = \arctan\frac{2AB}{1 - A^2 + B^2} \quad . \quad . \quad . \quad . \quad (8)$$

With the particular choice of terminal planes mentioned, α and β are no longer independent for a given junction since only one of the terminal planes is arbitrary. They are related, as also are A and B.

However, $2A/(1 + A^2 + B^2) = (1 - R^2)/(1 + R^2)$ is a constant since R is constant for a given junction.

(2.3) The Relation Between α, R, β and the Admittance Transformation Constants A and B for Junctions in Cascade

It is now possible to write down the relations between R, R_1, R_2, etc., in terms of the appropriate A and B values. We find from eqns. (2) and (6), (3) and (7), (4) and (8) that

$$\frac{(1 - R_1^2)(1 - R_2^2)}{(1 + R_1^2)(1 + R_2^2) + 4R_1R_2\cos\theta} = \frac{\cos\gamma_1\cos\gamma_2}{1 + \sin\gamma_1\sin\gamma_2\cos\theta} = \frac{1}{a} = \frac{2A}{1 + A^2 + B^2} \quad (9)$$

$$\frac{R_1(1 - R_2^2)\sin\theta}{R_2(1 + R_1^2) + R_1(1 + R_2^2)\cos\theta} = \frac{\sin\gamma_1\cos\gamma_2\sin\theta}{\sin\gamma_2 + \sin\gamma_1\cos\theta} = \frac{1}{b} = \tan(\beta_2 - \beta) \quad (10)$$

$$\frac{R_2(1 - R_1^2)\sin\theta}{R_1(1 + R_2^2) + R_2(1 + R_1^2)\cos\theta} = \frac{\cos\gamma_1\sin\gamma_2\sin\theta}{\sin\gamma_1 + \sin\gamma_2\cos\theta} = \frac{1}{c} = \tan(\alpha_1 - \alpha) \quad (11)$$

where

$$\tan\alpha = \frac{-2B}{1 - A^2 - B^2} \quad . \quad . \quad . \quad . \quad (12)$$

$$\tan\beta = \frac{2AB}{1 - A^2 + B^2} \quad . \quad . \quad . \quad . \quad (13)$$

$$\tan\alpha_1 = \frac{-2B_1}{1 - A_1^2 - B_1^2} \quad . \quad . \quad . \quad . \quad (14)$$

$$\tan\beta_2 = \frac{2A_2B_2}{1 - A_2^2 + B_2^2} \quad . \quad . \quad . \quad . \quad (15)$$

In eqns. (9), (10), and (11), the reciprocals of a, b, and c define either the left-hand or right-hand sides of these equations, and a new quantity γ has been introduced to simplify subsequent algebra, where

$$R_1 = \tan\frac{\gamma_1}{2},\ R_2 = \tan\frac{\gamma_2}{2} \quad . \quad . \quad . \quad (16)$$

We note that $\sin\gamma_1$, $\cos\gamma_1$, etc., are always positive quantities.

In our problem of setting up the matching section we require solutions of eqns. (9), (10), and (11) for R_1 and θ in terms of A_2, B_2 and A, B if the matching section precedes the junction, or for R_2 and θ in terms of A_1, B_1 and A, B if the junction precedes the matching section. For example, in the first instance, A_2 and B_2 specify the junction properties, whilst A and B specify the admittance transformation required. The solutions for R_1 and θ are

$$\left[1+\frac{a^2}{b^2}+(a^2-1)\cos^2\gamma_2\right]\cos^2\gamma_1 - 2a\cos\gamma_2\left(1+\frac{1}{b^2}\right)\cos\gamma_1+\cos^2\gamma_2\left(1+\frac{1}{b^2}\right)=0 \quad (17)$$

where $$\frac{1}{a}=\frac{2A}{1+A^2+B^2},\ \frac{1}{b}=\tan(\beta_2-\beta) \quad . \quad . \quad (18)$$

and $$\sin\theta=\frac{a\cos\gamma_1-\cos\gamma_2}{b\sin\gamma_1\sin\gamma_2}=\frac{a\cos\gamma_2-\cos\gamma_1}{c\sin\gamma_1\sin\gamma_2} \quad (19)$$

$$\cos\theta=\frac{a\cos\gamma_1\cos\gamma_2-1}{\sin\gamma_1\sin\gamma_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Those for R_2 and θ are

$$\left[1+\frac{a^2}{c^2}+(a^2-1)\cos^2\gamma_1\right]\cos^2\gamma_2 - 2\cos\gamma_1\left(1+\frac{1}{c^2}\right)\cos\gamma_2+\cos^2\gamma_1\left(1+\frac{1}{c^2}\right)=0 \quad (21)$$

where $$\frac{1}{a}=\frac{2A}{1+A^2+B^2} \quad . \quad . \quad . \quad . \quad . \quad (22)$$

as before, and $1/c=\tan(\alpha_1-\alpha)$

with $$\sin\theta=\frac{a\cos\gamma_2-\cos\gamma_1}{c\sin\gamma_1\sin\gamma_2} \quad . \quad . \quad . \quad (23)$$

Incidentally, eliminating θ we find a relation between a, b and c as follows:

$$c=\frac{b(\cos\gamma_1-a\cos\gamma_2)}{\cos\gamma_2-a\cos\gamma_1} \quad . \quad . \quad . \quad . \quad (24)$$

whilst, from eqn. (20),

$$\frac{da}{d\theta}=-\tan\gamma_1\tan\gamma_2\sin\theta \quad . \quad . \quad . \quad (25)$$

This equation is useful since it is necessary, as described later, to establish the value of θ with no additional phase shift introduced between the reactor and junction.

3. CALCULATION OF MATCHING SECTION PARAMETERS

In view of the number of equations involved, it is worth while outlining the calculations, since these depend upon whether the input or output terminal planes are fixed, and whether the matching section precedes the junction, or vice versa.

(3.1) Admittance Transformation and Position of Output Terminal Plane Specified

We assume that the matching section precedes the junction and that an admittance transformation $y'=Ay''+jB$ is specified. The constants A_2 and B_2 of the junction corresponding to the position of the chosen output plane are found by experiment as described in Section 2.2, with the reactor adjusted to produce no reflection. Cos γ_1 is obtained as a solution of the quadratic equation (17), in which a and b are computed from A, B and A_2, B_2, and θ is obtained from eqns. (19) and (20). The reactor and phase shifter are then adjusted in accordance with a previous calibration, as discussed in Section 4, to give the calculated reflection and phase shift; the admittance transformation then obtained can be checked by measurement. Two points may be noted.

First, before the required phase shift can be introduced it is necessary to find the value of θ, say θ_0, with the phase shifter set to a known value. Secondly, it is necessary to find which of the two calculated values of R_1, and hence θ, give the required admittance transformation. θ_0 can be found by adjusting the reactor to produce a known reflection (it is convenient to make R_1 equal to R_2) and then to find A and B by experiment. A_2 and B_2 will be known from previous measurement, so that θ_0 can be found from eqns. (19) and (20) without ambiguity. The proper value of R_1 is most easily decided by experiment, i.e. by measurement of A and B with the two calculated values of reflection and phase shift, as discussed in Section 4; it can also be decided by considering the admittance transformations of the equivalent circuit in the Smith chart.

(3.2) Position of the Input Terminal Plane Specified

In this case, we assume that the junction precedes the matching section, and that the position of the input terminal plane and the admittance transformation are specified. A_1 and B_1 are found by experiment with the matching section non-reflecting, and thus θ can be calculated from eqns. (19) and (20), and R_2 from eqn. (21). The comment in Section 3.1 concerning the ambiguity in these values is also applicable here, and it is necessary, of course, to find the value of θ_0.

(4) EXPERIMENTAL RESULTS

In order to verify the theory given in Section 2, experiments were carried out using a waveguide system consisting of a movable short-circuit, a variable reactor, a phase shifter, and a second variable reactor preceded by a standing-wave detector, all connected in cascade. The waveguide was rectangular, size No. 10, and the wavelength was approximately 10 cm. The short-circuit could be replaced by a matched termination to permit calibration of the reactors. Each reactor consisted of a rod of Distrene penetrating the narrow side of the waveguide in a transverse plane. The phase shift introduced by a reactor, i.e. the changes in the equivalent circuit parameters α and β, were small and could be allowed for by an adjustment of the phase shifter in the experiments.

The Distrene rods were each set to produce a voltage reflection coefficient of 0·237 when terminated with a matched load; then, choosing an arbitrary output terminal plane, the admittance transformation parameters A and B were found by using the movable short-circuit, the phase shift between the reactors being increased in steps of 20°. The values of A and B were plotted against phase shift, as shown in Fig. 2, and the input v.s.w.r. calculated from them was compared with that measured when using the standing-wave detector and a matched termination. The plot shows the measure of agreement obtained.

Next, with the Distrene rod furthest from the short-circuit withdrawn, A_2 and B_2 were found by displacing the short-circuit $\lambda_{g/8}$ and $\lambda_{g/4}$. Their values were 1·465 and 0·355 respectively. Thus, from eqn. (6),

$$R_2=\sqrt{\frac{(1-1\cdot465)^2+0\cdot355^2}{(1+1\cdot465)^2+0\cdot355^2}}=0\cdot237$$

giving a v.s.w.r. of $(1+0\cdot237)/(1-0\cdot237)=1\cdot63$. The v.s.w.r. measured directly was 1·65.

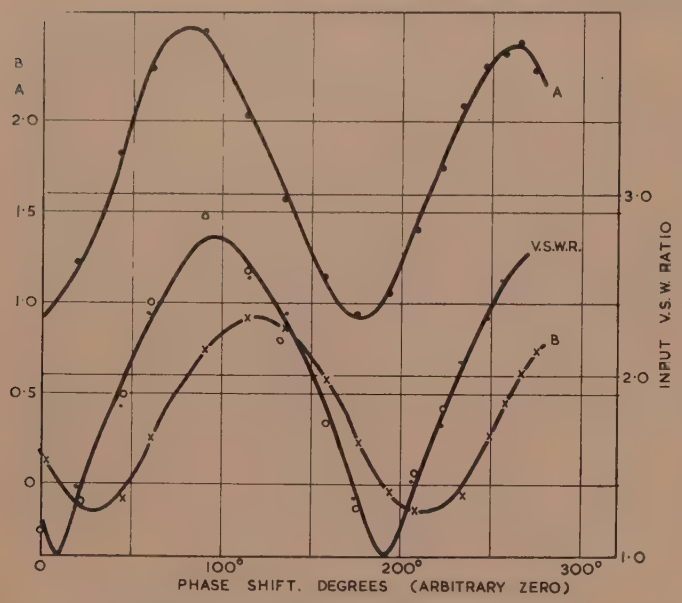


Fig. 2.—Variation of junction parameters A, B, and input-voltage standing-wave ratio with phase shift.

○ ○ ○ Measured v.s.w.r. using standing-wave detector.
● ● ● Calculated v.s.w.r. using the admittance transformation parameters.

The equivalent circuit parameter β_2 calculated from eqn. (15) was

$$\beta_2 = \text{arc tan}\frac{2 \times 1{\cdot}465 \times 0{\cdot}355}{1 - 1{\cdot}465^2 + 0{\cdot}355^2} = \text{arc tan}\,(-1{\cdot}015)$$

giving $\beta_2 = -45°$ approximately.

Now choosing from Fig. 2 a pair of values $A = 1{\cdot}5$ and $B = 0{\cdot}86$ at a phase of 136°,

$$\beta = \text{arc tan}\frac{2 \times 1{\cdot}5 \times 0{\cdot}86}{1 - 1{\cdot}5^2 + 0{\cdot}86^2} = \text{arc tan}\,(-5{\cdot}16)$$

or $\beta = -79°3'$

Thus, from eqn. (10),

$$\frac{1}{b} = \tan\,(-45° + 79°3') = 0{\cdot}675$$

and from eqn. (9),

$$\frac{1}{a} = \frac{2 \times 1{\cdot}5}{1 + 1{\cdot}5^2 + 0{\cdot}86^2} = 0{\cdot}75$$

R_2 found by a previous measurement was 0·237, so that, from eqn. (16),

$$\gamma_2 = 26°40',\ \cos\gamma_2 = 0{\cdot}894,\ \sin\gamma_2 = 0{\cdot}449$$

which gives the quadratic equation (17) in $\cos\gamma_1$ as follows:

$$2{\cdot}43\cos^2\gamma_1 - 3{\cdot}47\cos\gamma_1 + 1{\cdot}16 = 0$$

The solutions are $\cos\gamma_1 = 0{\cdot}896$, or 0·534. Referring to eqn. (19) the value $\cos\gamma_1 = 0{\cdot}534$ gives $\sin\theta$ negative, whereas from Fig. 2, $dS/d\theta$ is negative, S being the v.s.w.r. It may be shown that $dS/d\theta$ and $da/d\theta$ always have the same sign. Hence from eqn. (25), $\sin\theta$ is positive. From eqn. (20), $\cos\gamma_1 = 0{\cdot}896$ gives

$$\cos\theta = \frac{1{\cdot}33 \times 0{\cdot}896 \times 0{\cdot}894 - 1}{0{\cdot}449 \times 0{\cdot}446} = 0{\cdot}356$$

i.e. $\theta = 69°4'$. Thus the correct value for $\cos\gamma_1$ is 0·896, giving $R_1 = \sqrt{(1 - 0{\cdot}896)/(1 + 0{\cdot}896)} = 0{\cdot}237$, which agrees with the measured value.

We have established a connection between θ and the scale of phase shift, and therefore to obtain a match [$\cos\theta = -1$ with $R_1 = R_2$ from eqn. (2)], i.e. $A = 1$ and $B = 0$, we note that the phase shift must be increased by $\frac{1}{2}(180° - 69°4') = 56°28'$, or decreased by $\frac{1}{2}(180° + 69°4') = 124°32'$. Referring to Fig. 2 we see that this condition is obtained at phases of +56° and −124°, relative to 136°, at which $A = 1{\cdot}5$ and $B = 0{\cdot}86$.

(5) CONCLUSIONS

The technique of setting up a matching section in cascade with a lossless junction to obtain a particular admittance transformation is useful when a short-circuit termination is applicable for measurement purposes. The admittance transformation is valid for terminal planes which are defined by the position of voltage zeros with the short-circuit termination. This limits the choice of position of terminal planes to one plane only. However, the restriction is unimportant when the junction is intended to transform the characteristic impedance of the waveguide at one side of the junction to a particular impedance value at the other side.

The experimental data required for setting up the matching section does not exceed that required by any other method, and whilst the calculation of data is relatively long, it is straightforward and only involves real numbers. A trial-and-error method of adjustment of the matching section is feasible, but the systematic method outlined has proved to be quicker.

(6) ACKNOWLEDGMENTS

The work was carried out under a contract placed by the Admiralty, to whom acknowledgment is made for permission to publish the paper. The author is also indebted to his colleagues for advice and assistance.

(7) REFERENCES

(1) RAGAN, G. L.: 'Microwave Transmission Circuits' (McGraw-Hill, 1948).
(2) BARLOW, H. M., and CULLEN, A. L.: 'Microwave Measurements' (Constable, 1950).
(3) KING, R. W. P.: 'Transmission Line Theory' (McGraw-Hill, 1955).
(4) DESCHAMPS, G. A.: 'Determination of Reflection Coefficients and Insertion Loss of a Waveguide Junction', *Journal of Applied Physics*, 1953, **24**, p. 1046.
(5) STORER, J. E., SHEINGOLD, L. S., and STEIN, S.: 'A Simple Graphical Analysis of a Two-Port Waveguide Junction', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 1004.
(6) DUKES, J. M. C.: 'Waveguides and Waveguide Junctions', *Wireless Engineer*, 1955, **32**, pp. 65 and 266.
(7) SHURMER, H. V.: 'Transformation of the Smith Chart through Lossless Junctions', *Proceedings I.E.E.*, Monograph No. 267 R, December 1957 (**105 C**, p. 177).
(8) SLATER, J. C.: 'Microwave Electronics' (McGraw-Hill, 1950).
(9) MONTGOMERY, C. G., DICKE, R. H., and PURCELL, E. M.: 'Principles of Microwave Circuits' (McGraw-Hill, 1948).

(8) APPENDICES

(8.1) Relation between Constants in Two Bilinear Transformations

The bilinear transformation relating the voltage reflection coefficients at the input and output planes of a lossless junction may be expressed in two forms as follows. Firstly,

$$r' = \varepsilon^{j\alpha}\frac{R + \varepsilon^{j\beta}r''}{1 + R\varepsilon^{j\beta}r''} \quad \ldots\ldots \quad (26)$$

or
$$r'' = \varepsilon^{-j\alpha}\frac{r' - R\varepsilon^{j\alpha}}{\varepsilon^{j\alpha} - Rr'} \quad . \quad . \quad . \quad . \quad (27)$$

where R is a real number equal to the magnitude of the input-voltage reflection coefficient with the junction terminated in a matched load, and α and β are phase constants which depend upon the position of the terminal planes.

Secondly,
$$r' = \frac{E + Dr''}{1 + Cr''} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

or
$$r'' = \frac{E - r'}{Cr' - D} \quad . \quad . \quad . \quad . \quad . \quad (29)$$

The relationship between these two sets of constants can be found from the identities (26) $\equiv$ (28) and (27) $\equiv$ (29) by giving r' and r'' special values.

Thus putting r'' and r' zero in turn, we find that $R\varepsilon^{j\alpha} = E$, $R\varepsilon^{-j\beta} = E/D$

Therefore, $\varepsilon^{j(\alpha+\beta)} = D$ and $R\varepsilon^{j\alpha} = E$

Putting $r' = 1$ and substituting for D and E in eqn. (28), we obtain $C = R\varepsilon^{j\beta}$. Finally, eliminating $\varepsilon^{j\beta}$ we find that

$$\left.\begin{aligned} R &= \sqrt{\left(\frac{EC}{D}\right)} \\ \alpha &= \tfrac{1}{2}\arg\frac{DE}{C} \\ \text{and} \quad \beta &= \tfrac{1}{2}\arg\frac{DC}{E} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (30)$$

(8.2) Transformation of Voltage Reflection Coefficient through Two Cascaded Junctions

From eqn. (26) the transformations of voltage reflection coefficient through the junctions are

$$r_1' = \varepsilon^{j\alpha_1}\frac{R_1 + \varepsilon^{j\beta_1}r_1''}{1 + R_1\varepsilon^{j\beta_1}r_1''} \quad . \quad . \quad . \quad . \quad (31)$$

and
$$r_2' = \varepsilon^{j\alpha_2}\frac{R_2 + \varepsilon^{j\beta_2}r_2''}{1 + R_2\varepsilon^{j\beta_2}r_2''} \quad . \quad . \quad . \quad . \quad (32)$$

The reflection coefficient r_1'' at the output plane of the first junction is equal to that at the input plane of the second junction, r_2', modified in phase only since both planes are in the common guide between junctions. Therefore $r_1'' = r_2'\varepsilon^{-4\pi l/\lambda_g}$, where l is the length of waveguide between the terminal planes. Substituting for r_1'' in eqn. (31) in terms of r_2' we find that the relation between r_1' and r_2'' is

$$r_1' = \varepsilon^{j\alpha_1}\left[\frac{\dfrac{R_1 + R_2\varepsilon^{-j\theta}}{1 + R_1R_2\varepsilon^{-j\theta}} + \varepsilon^{j\beta_2}\dfrac{R_1R_2 + \varepsilon^{-j\theta}}{1 + R_1R_2\varepsilon^{-j\theta}}r_2''}{1 + \varepsilon^{j\beta_2}\dfrac{R_2 + R_1\varepsilon^{-j\theta}}{1 + R_1R_2\varepsilon^{-j\theta}}r_2''}\right] \quad (33)$$

where
$$\theta = \frac{4\pi l}{\lambda_g} - \beta_1 - \alpha_2$$

Eqn. (33) can be put in the form of eqn. (26) using the relations of eqn. (30), and we find that

$$R^2 = \frac{(R_2 + R_1\varepsilon^{-j\theta})(R_1 + R_2\varepsilon^{-j\theta})}{(1 + R_1R_2\varepsilon^{-j\theta})(R_1R_2 + \varepsilon^{-j\theta})} \quad . \quad . \quad . \quad (34)$$

$$\varepsilon^{j2\alpha} = \varepsilon^{j2\alpha_1}\frac{(R_1 + R_2\varepsilon^{-j\theta})(R_1R_2 + \varepsilon^{-j\theta})}{(1 + R_1R_2\varepsilon^{-j\theta})(R_2 + R_1\varepsilon^{-j\theta})} \quad . \quad . \quad (35)$$

$$\varepsilon^{j2\beta} = \varepsilon^{j2\beta_2}\frac{(R_2 + R_1\varepsilon^{-j\theta})(R_1R_2 + \varepsilon^{-j\theta})}{(1 + R_1R_2\varepsilon^{-j\theta})(R_1 + R\varepsilon^{-j\theta})} \quad . \quad . \quad (36)$$

On simplification, eqns. (34), (35), and (36) give eqns. (2), (3), and (4) respectively.

621.372

The Institution of Electrical Engineers
Paper No. 2771 E
Mar. 1959
©

THE POWER RADIATED BY A SURFACE WAVE CIRCULATING AROUND A CYLINDRICAL SURFACE

By Professor H. E. M. BARLOW, Ph.D., B.Sc.(Eng.), Member.

(*The paper was first received 9th May, and in revised form 21st September*, 1958.)

SUMMARY

Recognizing the evanescent character of the surface-wave field distribution over the equi-phase planes and the important part played by the inclination of these planes with the normal to the interface when power is transferred across it, calculations are made for radiation arising when a wave of this kind circulates around a highly reactive supporting surface of cylindrical form. It is concluded that, when the surface has a finite loss, there will be a particular radius of curvature for which the surface wave progresses for a limited distance without attenuation.

(1) INTRODUCTION

Surface waves are characterized by an evanescent field structure over the wavefront, and, for the particular case of the Zenneck wave supported by a flat surface forming the junction of two homogeneous media, the wave structure is represented by an exponential decay of the field on both sides of the interface. We are usually concerned with air as one medium surrounding another of finite loss which we call the waveguide, and since there can be no radiation in these conditions, the only flow of energy across the supporting surface is into the guide, at a rate sufficient to supply the losses.

In an earlier paper, Cullen and the author[1] drew attention to the behaviour of this Zenneck form of wave when circulating around a curved cylindrical surface and pointed out that radiation occurs even when the radius of curvature is very large and the surface highly reactive. By analogy with a cut-off waveguide a simple physical interpretation of the radiation process was given, showing how the adjacent equi-phase planes, diverging as they do with increasing distance from the curved surface, cause the evanescent field to be transformed into a propagating one when the distance between two such planes becomes greater than half the free-space wavelength (Fig. 1). That conception will now be developed further and used to provide a basis for the calculation of the power radiated in these circumstances. It will be apparent that, when radiation occurs as a result of the curvature of the supporting surface, the roles of the two media must, in a sense, be interchanged because energy then passes across the interface outwards into the surrounding air. Thus, if the guide is to be capable of performing its allotted task, it must have a highly reactive surface storing the energy of a trapped wave from which leakage occurs as the wave progresses.

Particularly relevant to this subject is the work of Elliott,[2] who examined, from a somewhat different point of view, the characteristics of circulating surface waves supported by corrugated and dielectric-coated cylindrical surfaces. Elliott's results lead to field distributions of the same form as those derived here, but he does not attempt to calculate the power radiated or to trace the physical processes involved which the present paper suggests.

(2) FIELD EQUATIONS

If, in the first place, we consider a highly reactive lossless cylindrical surface (Fig. 1) forming the waveguide surrounded by air and imagine a Zenneck form of wave to be circulating around its circumference, we shall expect to be able to satisfy

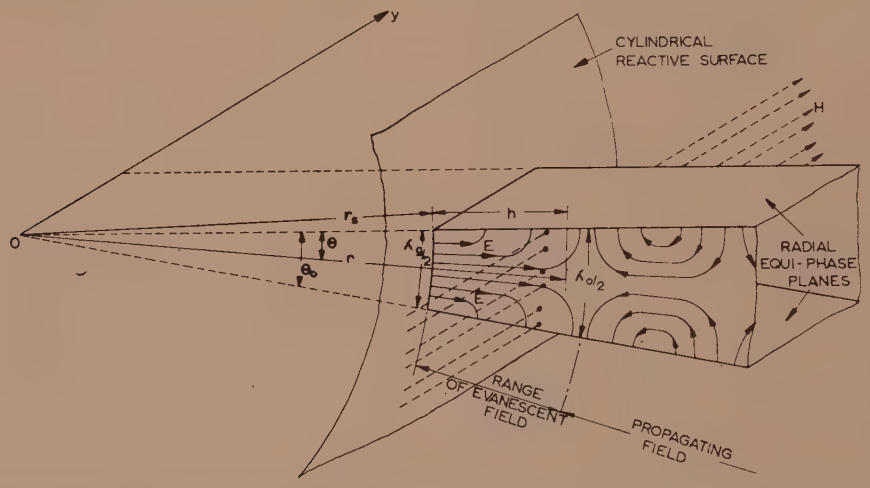


Fig. 1.—Co-ordinate system and radial non-radiating field pattern outside the supporting surface.

Maxwell's equations in cylindrical co-ordinates in terms of the three field components E_r, E_θ and H_y, there being no change in the y-direction. Assuming the usual sinusoidal variation in time and using script letters for the field components dependent upon space co-ordinates only, we find the wave equation

$$\frac{\partial^2 \mathscr{H}_y}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 \mathscr{H}_y}{\partial \theta^2} + \frac{1}{r}\frac{\partial \mathscr{H}_y}{\partial r} = \kappa^2 \mathscr{H}_y \quad . \quad . \quad (1)$$

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Prof. Barlow is Pender Professor of Electrical Engineering at University College, London.

where $$\kappa^2 = j\omega\mu(\sigma + j\omega\epsilon) \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Now $$\mathscr{H}_y = Af_1(\theta)f_2(r) \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and if, for the present, we suppose that equal and oppositely directed circulating waves are set up giving a stationary wave pattern so that the radial equi-phase planes remain in fixed positions,

$$f_1(\theta) = \cos(\beta r\theta) \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where βr is constant, and the wave equation becomes

$$\frac{\partial^2 f_2}{\partial r^2} + \frac{1}{r}\frac{\partial f_2}{\partial r} - u^2 f_2 = 0 \quad . \quad . \quad . \quad . \quad (5)$$

with $$u^2 = \kappa^2 + \beta^2 \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Since the medium outside the supporting surface is air,

$$\kappa^2 = -\omega^2\mu_0\epsilon_0$$

and with βr constant, it is clear from eqn. (6) that u must be a function of r.

The only way in which u could be kept constant would be by using an inhomogeneous dielectric, and in that case eqn. (1) would have to be replaced by

$$\frac{\partial^2 \mathscr{H}_y}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 \mathscr{H}_y}{\partial \theta^2} + \left[\frac{1}{r} - \frac{\epsilon'(r)}{\epsilon(r)}\right]\frac{\partial \mathscr{H}_y}{\partial r} = \kappa^2 \mathscr{H}_y \quad . \quad (7)$$

to which no simple solution appears to be available.

However, we can find a solution to eqn. (5) for the homogeneous air dielectric surrounding the guide, which we now wish to consider. Thus, to satisfy the boundary conditions at adjacent equi-phase planes separated by the angle θ_0, we must have

$$\beta r = \pi/\theta_0 = \nu \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and putting $$T^2 = -\kappa^2 = \omega^2\mu_0\epsilon_0 = \left(\frac{2\pi}{\lambda_0}\right)^2 \quad . \quad . \quad (9)$$

we can rewrite eqn. (5) as

$$\frac{\partial^2 f_2}{\partial r^2} + \frac{1}{r}\frac{\partial f_2}{\partial r} + \left(T^2 - \frac{\nu^2}{r^2}\right)f_2 = 0 \quad . \quad . \quad . \quad (10)$$

to which the appropriate general solution is

$$f_2 = B_1 \mathrm{H}_\nu^{(2)}(Tr) + B_2 \mathrm{H}_\nu^{(1)}(Tr) \quad . \quad . \quad . \quad . \quad (11)$$

To retain the character of a surface wave, the field in the region immediately outside the highly reactive surface must have an evanescent distribution in the radial direction. Tr is less than ν by perhaps 5%, while ν is of the order of 100 upwards, so that it is legitimate to apply the Liouville approximations to the Hankel functions (see Appendix 9). We find that $B_1\mathrm{H}_\nu^{(2)}(Tr)$ represents the 'incident' field decreasing as r increases, whilst $B_2\mathrm{H}_\nu^{(1)}(Tr)$ represents the 'reflected' field increasing with r. Similarly for large arguments, when Tr is greater than ν, these quantities represent respectively 'incident' and 'reflected' waves. The retention of the second term in eqn. (11) implies, of course, a physical discontinuity at a finite radius.

The field components of the stationary surface wave are therefore

$$\mathscr{H}_y = \mathscr{H}_y^+ + \mathscr{H}_y^- = \cos(\nu\theta)[C_1\mathrm{H}_\nu^{(2)}(Tr) + C_2\mathrm{H}_\nu^{(1)}(Tr)] \quad (12)$$

$$\mathscr{E}_\theta = \mathscr{E}_\theta^+ + \mathscr{E}_\theta^- = \frac{j}{\omega\epsilon}\cos(\nu\theta)\left\{C_1\left[\frac{\nu}{r}\mathrm{H}_\nu^{(2)}(Tr) - T\mathrm{H}_{\nu+1}^{(2)}(Tr)\right] + C_2\left[-\frac{\nu}{r}\mathrm{H}_\nu^{(1)}(Tr) - T\mathrm{H}_{\nu+1}^{(1)}(Tr)\right]\right\} \quad (13)$$

$$\mathscr{E}_r = \mathscr{E}_r^+ + \mathscr{E}_r^- = \frac{j\beta}{\omega\epsilon}\sin(\nu\theta)[C_1\mathrm{H}_\nu^{(2)}(Tr) + C_2\mathrm{H}_\nu^{(1)}(Tr)] \quad (14)$$

Incidentally, when $r \to \infty$, eqn. (10) reduces to

$$\frac{\partial^2 f_2}{\partial r^2} + T^2 f_2 = 0$$

giving $f_2 = Be^{-ur}$ in accordance with the requirements for a flat surface.

The magnetic-field component corresponding to eqn. (12), but for a circulating wave of the same amplitude and moving in the $+\theta$-direction, is

$$\mathscr{H}_y = \mathscr{H}_y^+ + \mathscr{H}_y^- = \varepsilon^{-j\nu\theta}[C_1\mathrm{H}_\nu^{(2)}(Tr) + C_2\mathrm{H}_\nu^{(1)}(Tr)] \quad (15)$$

If the medium below the supporting surface had a finite loss, $\nu = \beta r$ would become complex and we should have to replace β^2 by $-\gamma^2$, where $\gamma = \alpha + j\beta$ and includes the attenuation coefficient α as well as the phase coefficient β.

Thus we should have to deal with Hankel functions of complex order, and Elliott[2] has used expansions of these functions in the form proposed by Olver[3] to calculate propagation coefficients for values of Tr_s ranging from 60 to 960. The result shows that, in normal circumstances, α is very small compared with β. In the analysis which follows, ν is regarded as purely real, and, for the purpose of calculating radiation losses, equivalent provision is made in terms of the tilt of the equi-phase planes.

(3) CALCULATION OF THE POWER RADIATED FROM A SURFACE WAVE SUPPORTED BY A HIGHLY REACTIVE LOSSLESS CYLINDRICAL SURFACE

(3.1) Cut-Off Waveguide Method

In the preceding Section it has been shown that Maxwell's equations can be satisfied by the Zenneck form of surface wave outside a cylindrical surface. The three field components, $\mathscr{H}_y$, $\mathscr{E}_\theta$ and $\mathscr{E}_r$, form a purely radial field pattern of evanescent structure in the neighbourhood of the surface. For the purpose of calculating the power radiated we can regard the field pattern as stationary in space and consider one section of it represented by a sector subtending an angle θ_0 and occupying the space between two adjacent equi-phase planes (Fig. 1). This forms what is equivalent to a wedge-shaped waveguide operating below cut-off in the vicinity of the surface. Power can flow radially outwards, within the sector, only by reason of a reflection taking place at some discontinuity remote from the surface. The mechanism envisaged is the same as that which permits power to flow along a cut-off waveguide.[4] However, in the case under consideration, there can be no radiation when the equi-phase planes are radial, and to obtain power flowing away from the surface these planes must be tilted backwards through an angle η as shown in Fig. 2(*a*). This angle may be positive or negative according to whether the power flows into or out of the surface, and if ψ_0 is the real part of the Brewster angle,

$$\eta = \pi/2 - \psi_0.$$

It is clear, therefore, that the effective reflection coefficient that makes possible radiation from the surface wave must be a function of η.

In order to produce a reflection there must be a physical discontinuity. With a purely radial field pattern there will be a transition from the evanescent distribution in the immediate neighbourhood of the surface to a propagating wave of the same structure further out, but this would not be expected to introduce any reflection. Moreover, the radial pattern ensures that, from all points around the circumference of the guide, the field combines to form an expanding cylindrical wave of the original distribution. On the other hand, as soon as a finite

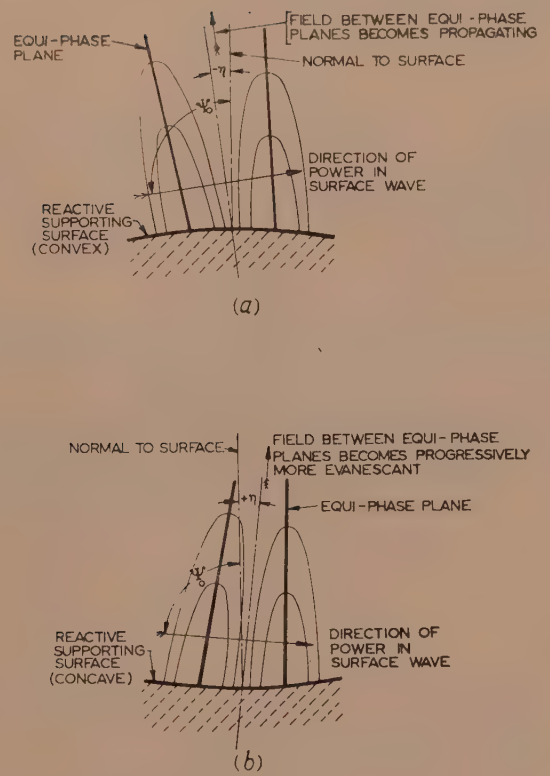


Fig. 2.—Field pattern of surface wave, supported by a cylindrical surface.

(*a*) Divergent wedge-shaped guide between adjacent equi-phase planes tilted backwards, giving radiation.
(*b*) Convergent wedge-shaped guide between adjacent equi-phase planes tilted forward, giving power into the surface.

angle $-\eta$ is introduced (for which power flows away from the surface), interference is set up between the fields associated with adjacent backward-tilted sectors bounded by the corresponding equi-phase planes, which, for the present analysis, are regarded as stationary in space. This interference will tend to break up the surface-wave field pattern, and in the worst case when $\eta = \pi/2$, it is reasonable to expect that the original pattern is transformed completely in free space to a plane wave. Thus we can write in terms of the reflection coefficient ρ_s at the surface,

$$\rho_s = \rho_0 \sin(-\eta) \simeq -\rho_0\eta \quad . \quad . \quad . \quad . \quad (16)$$

for small angles,

where ρ_0 is the reflection coefficient arising from the complete transformation of the surface wave to a plane wave.

It is important to recognize that, on the basis of this argument, any divergence or convergence of the equi-phase planes has nothing whatever to do with radiation and consequently is not, in itself, a factor affecting the value of ρ_0. To be consistent in calculating this quantity we must therefore exclude any effect of curvature of the surface.

Having thus made provision for the tilt of the field pattern, we can now disregard this and proceed to calculate the power that would be expected to flow outwards within the cut-off sector guide, with the reflection coefficient ρ at any radius r. The field impedance looking radially outwards from the surface is

$$Z^+ = \frac{\mathscr{E}_\theta^+}{\mathscr{H}_y^+} = \frac{j}{\omega\epsilon}\left[\frac{\nu}{r} - T\frac{H_{\nu+1}^{(2)}(Tr)}{H_\nu^{(2)}(Tr)}\right] \quad . \quad . \quad (17)$$

and in the region near the surface where $Tr < \nu$ we can apply the Liouville approximations to the Hankel functions yielding

$$\frac{H_{\nu+1}^{(2)}(Tr)}{H_\nu^{(2)}(Tr)} \simeq \left(\frac{\varepsilon\nu}{Tr}\right)(1+M)\varepsilon^{-M} \simeq \frac{\varepsilon\nu}{Tr}$$

where $M = \sqrt{[1-(Tr/\nu)^2]}$ and use is made of the further approximation $\varepsilon^M \simeq 1 + M$

$$\text{so that} \quad Z^+ = \frac{j\nu}{\omega\epsilon r}(1-\varepsilon) = -j1{\cdot}72\left(\frac{\beta}{\omega\epsilon}\right) = jX^+ \quad (18)$$

This is a purely capacitive reactance as required for an evanescent field.

In the same way, when looking into the surface, we find a purely inductive reactance (of the same value) according to the requirements for the support of a surface wave at an interface between lossless media and no radiation. It is, of course, the reflected field from outer space that produces the radiation, and when this field is superimposed on the incident one, the surface impedance of the guide acquires a negative-resistance component to provide for the power flow out of the surface. There is nothing inconsistent in this. It simply means that the energy associated with the wave is mostly stored within the surface, the effect of radiation being to give to the surrounding medium equivalent losses which must then be supplied by energy crossing the interface.

From the electric and magnetic field components tangential to the surface, namely $\mathscr{E}_\theta$ and $\mathscr{H}_y$, we can calculate[5] the radial power density p as follows.

Dropping the direction subscripts, we have

$$p = \mathscr{R}[(\mathscr{E}^+ + \mathscr{E}^-)(\mathscr{H}_*^+ + \mathscr{H}_*^-)] \quad . \quad . \quad (19)$$

where $\mathscr{H}_*^+$ and $\mathscr{H}_*^-$ are, respectively, the complex conjugates of the incident and reflected magnetic field components. Since the field impedance is purely reactive $\mathscr{E}^+$ must be in time quadrature with $\mathscr{H}^+$, and $\mathscr{E}^-$ in time quadrature with $\mathscr{H}^-$, so that

$$p = \mathscr{R}(\mathscr{E}^+\mathscr{H}_*^- + \mathscr{E}^-\mathscr{H}_*^+) \quad . \quad . \quad . \quad (20)$$

$$\text{but} \quad \rho = |\rho|\varepsilon^{j\phi} = \frac{\mathscr{E}^-}{\mathscr{E}^+} = -\frac{\mathscr{H}^-}{\mathscr{H}^+} \quad . \quad . \quad . \quad (21)$$

and using eqn. (18) we find

$$p = -2|\rho|X^+|\mathscr{H}^+|^2 \sin\phi \quad . \quad . \quad . \quad . \quad (22)$$

Inserting the values given by eqn. (12) for $\mathscr{H}^-$ and $\mathscr{H}^+$ in eqn. (21) and applying the Liouville approximations to the Hankel functions together with $\varepsilon^M \simeq 1 + M$ where $M = \sqrt{[1-(Tr/\nu)^2]}$ and $\nu! \simeq \nu^\nu\varepsilon^{-\nu}\sqrt{2\pi\nu}$ which, for $\nu > 10$, as in the present application, leads to negligible error (see Appendix 9), we get

$$\rho = \frac{C_2}{C_1}\left(\frac{Tr}{\nu}\right)^{2\nu}\varepsilon^{j\pi/2} \quad . \quad . \quad . \quad . \quad (23)$$

$$\text{Suppose} \quad C_1 = |C_1| \text{ and is purely real,}$$

$$\text{whilst} \quad \frac{C_2}{C_1} = \rho_1 = |\rho_1|\varepsilon^{j\theta} \quad . \quad . \quad . \quad . \quad . \quad (24)$$

$$\text{Then} \quad p = 1{\cdot}72\left[\frac{4|\rho_1|C_1^2\cos^2(\nu\theta)}{M\pi\omega\epsilon r}\right]\sin\phi \quad . \quad . \quad (25)$$

$$\text{where} \quad \phi = \phi_1 + \pi/2 \quad . \quad . \quad . \quad . \quad . \quad (26)$$

The total power P over a surface area represented by the arc $r\theta_0$ and unit length in the y-direction is therefore

$$P = 1{\cdot}72\left[\frac{2|\rho_1|\theta_0C_1^2}{M\pi\omega\epsilon}\right]\sin\phi \quad . \quad . \quad . \quad . \quad (27)$$

This power is directly proportional to θ_0 and (neglecting the very small variation of M with r) is independent of r, as would be expected.

It is of interest to observe that this form of radial evanescent field distribution offers a new type of cut-off attenuator in which the field varies along the length according to a Hankel function and there is a possible additional variable θ_0. Within the evanescent range, either the small argument or the Liouville approximations are applicable to the Hankel functions, and in the latter case, for the incident magnetic field we should have

$$\mathscr{H}^+ = jC_1\left(\frac{\nu}{Tr}\right)^{\nu}\sqrt{\frac{2}{\pi\nu M}}\cos(\nu\theta) \quad . \quad . \quad . \quad (28)$$

Pursuing the calculation of power radiated by the surface wave, it is convenient to rewrite eqn. (27) in terms of the reflection coefficient $\rho_s = |\rho_s|\varepsilon^{j\phi_s}$ at the surface where $r = r_s$.

Thus
$$|\rho_s| = |\rho_1|\left(\frac{Tr_s}{\nu}\right)^{2\nu} \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

and
$$\phi = \phi_s = (\phi_1 + \pi/2) \quad . \quad . \quad . \quad . \quad . \quad (30)$$

It will be noticed that the phase of the reflection coefficient remains unchanged over the range of the evanescent field.

Hence
$$P = 1{\cdot}72\left[\frac{2|\rho_s|\theta_0 C_1^2}{M\pi\omega\epsilon}\right]\left(\frac{\nu}{Tr_s}\right)^{2\nu}\sin\phi_s \quad . \quad . \quad (31)$$

or using eqn. (16),

$$P = -1{\cdot}72\left[\frac{2\eta|\rho_0|\theta_0 C_1^2}{M\pi\omega\epsilon}\right]\left(\frac{\nu}{Tr_s}\right)^{2\nu}\sin\phi_0 \quad . \quad . \quad (32)$$

where
$$\rho_0 = |\rho_0|\varepsilon^{j\phi_0} \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Bearing in mind that ρ_0 represents the reflection coefficient which arises from complete transformation of a pure surface wave to a free-space plane wave we can calculate its value from the relation

$$\rho_0 = \frac{Z - Z_0^+}{Z + Z_0^+} \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

with $Z = \sqrt{(\mu_0/\epsilon_0)}$ = Free-space wave impedance and

$$Z_0^+ = -j\frac{\beta}{\omega\epsilon} = -j\sqrt{\frac{\mu_0}{\epsilon_0}}$$

Z_0^+ is the wave impedance looking outwards from the corresponding flat surface and therefore deliberately excludes the effect of curvature which shows itself in Z^+ as given by eqn. (18). Power passing into or out of the surface is always associated with a tilt forwards or backwards of the equi-phase planes. This behaviour is in evidence in the case of a flat surface with a lossy medium on one or the other side of the interface. Divergence of these equi-phase planes, which arises from curvature, does not in any way affect the issue. Provision is made quite independently for the effect of curvature in its tendency to tilt the equi-phase planes. This is part of the general mechanism of energy flow across the interface, depending, as it does, solely on the net tilt which takes into account both absorption in the media concerned and curvature of the surface.

Eqn. (34) thus yields $\rho_0 = \varepsilon^{j\pi/2}$

so that $|\rho_0| = 1$ and $\phi_0 = \pi/2$

and inserting these values in eqn. (32), remembering that $\beta_s r_s$ we get

$$P = -1{\cdot}72\frac{2\eta\theta_0 C_1^2}{M\pi\omega\epsilon}\left(\frac{\beta_s}{T}\right)^{2\nu} \quad . \quad . \quad . \quad . \quad (35)$$

(3.2) Complex Brewster-Angle Method

We have seen that the equi-phase planes of the circulating surface wave are, in general, inclined at an angle $\pm\eta$ to the radial direction. Since the field pattern is maintained with a backward tilt (η negative) when radiation occurs, we can represent the conditions by a change of the angle θ to θ' (Fig. 3), such that

$$\theta' = \theta - \frac{h}{r}\eta \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

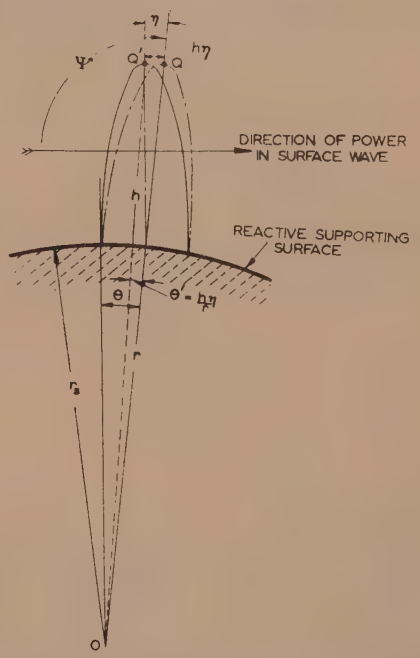


Fig. 3.—Change of angular co-ordinate accompanying radiation.

where the distance from the surface is

$$h = (r - r_s) \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

For the magnetic-field component of the equivalent circulating wave travelling in the $+\theta'$ direction, we have

$$\mathscr{H}_y = C_1\varepsilon^{-j\nu\theta'}\mathrm{H}^{(2)}(Tr) \quad . \quad . \quad . \quad . \quad (38)$$

Now
$$\eta = (\pi/2 - \psi)$$

where ψ is measured from the normal to the surface and corresponds to the Brewster angle.

In dealing with radiation conditions, we have to consider what is equivalent to a lossy medium, so that ψ becomes complex and

$$\psi = \psi_0 - j\chi \quad . \quad . \quad . \quad . \quad . \quad (39)$$

The direction of the power flow is determined by the real part of the Brewster angle ψ_0. Thus from eqns. (36), (37), (38) and (39), we get

$$\mathscr{H}_y = [C_1\mathrm{H}_\nu^{(2)}(Tr)][\varepsilon^{-j\nu\theta}\varepsilon^{-qh/r}] \quad . \quad . \quad . \quad (40)$$

where
$$q = \nu\chi - j\nu\left(\frac{\pi}{2} - \psi_0\right) = \nu\chi - j\nu\eta \quad . \quad . \quad (41)$$

Since h/r increases as r increases, it is clear that eqn. (40) represents a circulating wave in the $+\theta$ direction with an exponential decay in the radial direction arising from the tilt, superimposed on the Hankel function distribution. There is also a wave moving towards the surface when η is positive and away from the surface when η is negative. This satisfies the requirements.

Corresponding to the magnetic field given by eqn. (40), we have an electric field

$$\mathscr{E}_\theta = -\frac{1}{j\omega\epsilon}\frac{\partial \mathscr{H}_y}{\partial r} = j\frac{C_1}{\omega\epsilon r}\left[\varepsilon^{-j\nu\theta}\varepsilon^{-qh/r}\right]$$
$$\left[\left(\nu - q\frac{r_s}{r}\right)H_\nu^{(2)}(Tr) - (Tr)H_{\nu+1}^{(2)}(Tr)\right] \quad (42)$$

Using the Liouville approximations for the Hankel functions and calculating the radial power density p from

$$p = \mathscr{R}(\mathscr{E}_\theta \mathscr{H}_y^*) \quad . \quad . \quad . \quad . \quad . \quad (43)$$

we get

$$p = -\frac{2\eta C_1^2 r_s}{M\pi\omega\epsilon r^2}\left(\frac{\nu}{Tr}\right)^{2\nu}\varepsilon^{-2\nu\chi h/r} \quad . \quad . \quad . \quad (44)$$

or the total power P passing through an area represented by an arc $r\theta_0$ and unit depth is

$$P = pr\theta_0 = -\frac{2\eta C_1^2 r_s\theta_0}{M\pi\omega\epsilon r}\left(\frac{\beta}{T}\right)^{2\nu}\varepsilon^{-2\beta\chi h} \quad . \quad . \quad (45)$$

or in terms of surface values where $h = 0$,

$$P = -\frac{2\eta\theta_0 C_1^2}{M\pi\omega\epsilon}\left(\frac{\beta_s}{T}\right)^{2\nu} \quad . \quad . \quad . \quad . \quad (35a)$$

which differs only slightly in the numerical factor from eqn. (35) previously deduced.

It will be observed that, in applying the Liouville expressions for the Hankel functions the rather rough approximation $\varepsilon^M \simeq (1 + M)$ has also been employed. The degree of accuracy in the numerical factors, achieved as a result of the different mathematical processes, may therefore not always be precisely the same. The intention is, however, to try to give an insight into the physical processes involved, and to make an estimate of their influence on the mechanism, rather than to attempt a precise analysis of the problem.

(4) CONDITION FOR WHICH NO POWER PASSES ACROSS THE SUPPORTING SURFACE

Suppose we have a surface wave in air supported by a flat surface, below which is a homogeneous medium of finite loss: we know that power passes across the interface, being supplied from the energy contained in the field above the surface [Fig. 4(*a*)]. If the real part of the Brewster angle which determines the direction of the power within the surface is ψ_1 and its imaginary part is χ_1,

$$\tan(\psi_1 - j\chi_1) = \frac{Z}{Z_1} = m - jn \quad . \quad . \quad . \quad (46)$$

where $Z = \sqrt{\dfrac{\mu_0}{\epsilon_0}}$ = Wave impedance of air outside surface

and $Z_1 = \sqrt{\dfrac{j\omega\mu_0}{\sigma_1 + j\omega\epsilon_1}}$ = Wave impedance of medium below the surface.

Since
$$\eta_1 = (\pi/2 - \psi_1) \quad . \quad . \quad . \quad . \quad . \quad (47)$$

we find

$$m = \frac{\sin 2\eta_1}{2(\cosh^2\chi_1 - \cos^2\eta_1)} = \left\{\frac{\epsilon_0/\epsilon_1[\sqrt{(1+w^2)}+1]}{2(1+w^2)}\right\}^{1/2} \quad (48)$$

and
$$n = \frac{\sinh 2\chi_1}{2(\cosh^2\chi_1 - \cos^2\eta_1)} = -\left\{\frac{\epsilon_0/\epsilon_1[\sqrt{(1+w^2)}-1]}{2(1+w^2)}\right\}^{1/2} \quad (49)$$

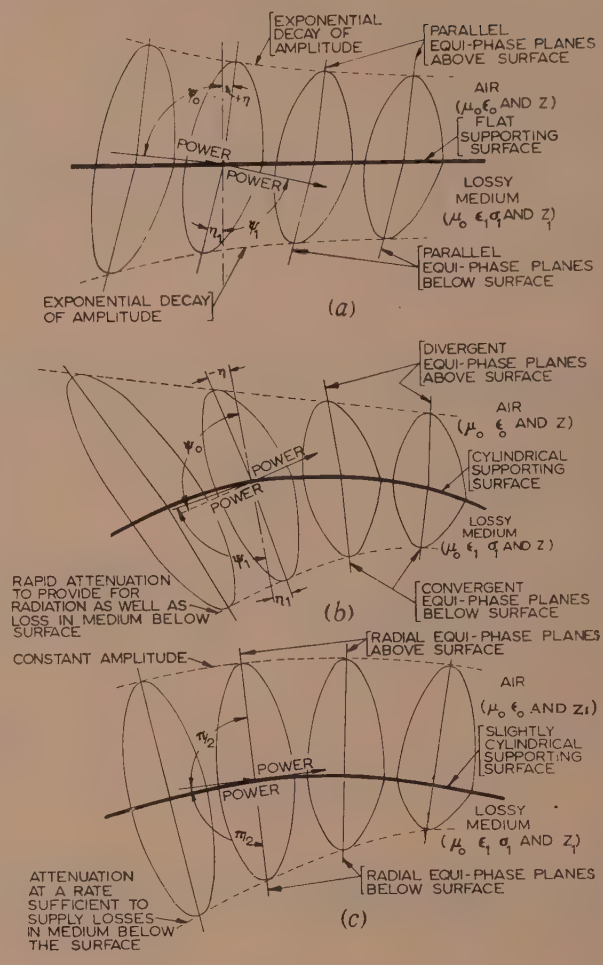


Fig. 4.—Electric-field patterns above and below the surface, for flat and cylindrical surfaces.
(*a*) Flat surface.
(*b*) Cylindrical surface of large radius.
(*c*) Cylindrical surface of very large radius.

where
$$w = \sigma_1/\omega\epsilon_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

Assuming η_1 to be a small angle,

$$m = \frac{\eta_1}{\sinh^2\chi_1} \text{ and } n = \frac{1}{\tanh\chi_1}$$

so that
$$\eta_1 = \left(\frac{m}{n^2 - 1}\right) \quad . \quad . \quad . \quad . \quad . \quad (51)$$

Using eqns. (48), (49) and (50) we can therefore calculate from eqn. (51) the value of η_1 for a medium below the surface, of given electrical characteristics.

As the surface is bent, radiation tends to arise and the equi-phase planes of the surface wave move from a forward to a backward tilt [Fig. 4(*b*)]. When the backward tilt angle $-\eta$ associated with the radiation process is equal to the forward tilt angle $+\eta_1$ associated with the losses in the surface, one would expect no power to cross the interface and the wave outside the surface to progress without attenuation [Fig. 4(*c*)]. This situation clearly cannot continue indefinitely but should be maintained until all the energy of the wave beneath the surface has been used up.

(5) POWER CARRIED BY THE SURFACE WAVE AROUND THE SURFACE

The power carried by the circulating surface wave around the surface per unit length in the y-direction is given by

$$P_\theta = \int_{r=r_s}^{r=\infty} \mathscr{R}[-\mathscr{E}_r \mathscr{H}_y^*]dr$$

To compute this quantity is clearly a difficult matter, partly because of the complexity of the expression involved but more particularly because of the strictly limited range of application of the Liouville approximations.

(6) CONCLUSIONS

Two methods have been developed for calculating the power radiated by a Zenneck form of surface wave circulating around a cylindrical surface, and they have been shown to lead to substantially the same result. It is concluded from the analysis that, for a particular curvature of the surface beneath which the medium has a finite loss, the surface wave should suffer no attenuation.

(7) ACKNOWLEDGMENTS

The author wishes to thank his colleague, Dr. John Brown, and Prof. A. L. Cullen for many stimulating discussions on surface-wave problems.

(8) REFERENCES

(1) BARLOW, H. E. M., and CULLEN, A. L.: 'Surface Waves', *Proceedings I.E.E.*, Paper No. 1482 R, April, 1953 (**100,** Part III, p. 329).
(2) ELLIOTT, R. S.: 'Azimuthal Surface Waves on Circular Cylinders', *Journal of Applied Physics*, 1955, **26,** p. 368.
(3) OLVER, F. W. J.: *Proceedings of the Cambridge Philosophical Society*, 1952, **48,** Part I, p. 414.
(4) BARLOW, H. E. M., and CULLEN, A. L.: 'Microwave Measurements' (Constable, 1950).
(5) CULLEN, A. L.: 'Waveguide Field Patterns in Evanescent Modes', *Wireless Engineer*, 1949, **26,** p. 317.

(9) APPENDIX

Liouville's Approximation to Hankel Functions

The Hankel functions as linear combinations of the Bessel functions $J_\nu(Tr)$ and $Y_\nu(Tr)$ can be evaluated within the range for which

$$\frac{(Tr)^2[(Tr)^2 + 4\nu^2]}{4[\nu^2 - (Tr)^2]^3} \ll 1 \quad . \quad . \quad . \quad . \quad (52)$$

from Liouville's approximations

$$J_\nu(Tr) = \frac{1}{\nu!}\left(\frac{Tr}{1+M}\right)^\nu \frac{\varepsilon^{\nu(M-1)}}{\sqrt{M}} \quad . \quad . \quad . \quad (53)$$

and

$$Y_\nu(Tr) = -\frac{(\nu-1)!}{\pi}\left(\frac{1+M}{Tr}\right)^\nu \frac{\varepsilon^{\nu(1-M)}}{\sqrt{M}} \quad . \quad . \quad (54)$$

where

$$M = \sqrt{[1 - (Tr/\nu)^2]} \quad . \quad . \quad . \quad . \quad (55)$$

We also have

$$\frac{Y_\nu(Tr)}{J_\nu(Tr)} = -\frac{(\nu-1)!(\nu)!}{\pi}\left(\frac{1+M}{Tr}\right)^{2\nu} \varepsilon^{2\nu(1-M)} \quad . \quad (56)$$

To ascertain whether these approximations are valid, we first calculate the ratio of the free-space wavelength λ_0 to the corresponding wavelength λ_g along the surface.[1] In the present application we must have a highly reactive surface, and for this computation we can therefore neglect losses in the surface and the effect of any slight curvature.

Thus

$$\frac{\lambda_0}{\lambda_g} = \sqrt{\left[1 + \left(\frac{X_s}{Z}\right)^2\right]} \quad . \quad . \quad . \quad . \quad (57)$$

where X_s is the surface reactance, being equal and opposite to X^+ in eqn. (18), and $Z = \sqrt{(\mu_0/\epsilon_0)}$ is the free-space wave impedance.

For the curved surface,

$$\nu = \beta r = \beta_s r_s \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and

$$T = 2\pi/\lambda_0 \quad . \quad . \quad . \quad . \quad . \quad (9)$$

so that

$$\nu/Tr = \left(\frac{\lambda_0}{\lambda_g}\right)\left(\frac{r_s}{r}\right) \quad . \quad . \quad . \quad . \quad (58)$$

Suppose we have a surface for which $X_s = 133$ ohms with a radius of curvature $r_s = 1$ m and consider a point distant 1 cm from the surface so that $r = 1{\cdot}01$ m. We find that $\nu = 222$, $Tr = 211$ and the condition of eqn. (52) is easily satisfied since the numerical value of the quantity on the left-hand side is $2{\cdot}78 \times 10^{-2}$.

From eqn. (55) we get in this case $M = 0{\cdot}3$, and putting

$$\nu! = \left(\frac{\nu}{\varepsilon}\right)^\nu \sqrt{(2\pi\nu)}$$

for these large values of ν eqn. (56) yields

$$\frac{Y_\nu(Tr)}{J_\nu(Tr)} = -160$$

In these circumstances, we therefore have

$$-Y_\nu(Tr) \gg J_\nu(Tr)$$

and it is justifiable to neglect the real part of the Hankel functions.

A further approximation which is applicable without serious error in evaluating the individual Hankel functions for the example given is

$$\varepsilon^M \simeq 1 + M \quad . \quad . \quad . \quad . \quad . \quad (59)$$

621.317.39 : 621.5

The Institution of Electrical Engineers
Paper No. 2758 M
Oct. 1958
©

THE RECOGNITION OF MOVING VEHICLES BY ELECTRONIC MEANS

By T. S. PICK, B.Sc.(Eng.), Member, and A. READMAN, Associate Member.

(*The paper was first received* 1*st February, and in revised form* 24*th July*, 1958. *It was published in October*, 1958, *and was read before* THE INSTITUTION 6*th November, the* SOUTH-WEST SCOTLAND SUB-CENTRE 11*th November, the* MERSEY AND NORTH WALES CENTRE 24*th November*, 1958, *the* SOUTH-EAST SCOTLAND SUB-CENTRE 17*th February, and the* EAST ANGLIAN SUB-CENTRE 2*nd March*, 1959.)

SUMMARY

The increasing congestion of traffic in London's streets has been a matter of major concern to London Transport for many years, and the maintenance of service regularity in these difficult circumstances has so far been achieved with some degree of success by additional supervision of operating personnel. The paper describes a method whereby, without any additions to the road staff, an increasing amount of information can be given at a central control point as to the movement of buses. This should enable measures to be taken before conditions deteriorate, and thus minimize service irregularities.

Details of an equipment which has been developed for this purpose are given, and the component items of the equipment, namely the identification markings on each bus, the roadside scanning and sending units and the central receiving and presentation cabinet, are described. A later Section of the paper forecasts improvements and developments which are visualized for the future, and although this equipment has been designed to deal with a specific bus operating problem, it should have numerous applications wherever a number of moving objects need to be recorded.

(1) INTRODUCTION

If one stands on the Embankment outside the Institution building to observe the running of buses, the sort of information obtained will be that, over a half-mile stretch of road, traffic is or is not moving normally, and also that for a given period the appropriate bus services are, in fact, operating. To check this last statement objectively, however, a written record with stop-watch timings would be necessary. Such a local record would be of little value unless it could be correlated with other local records along each of the routes and transmitted to someone in authority, who could direct remedial measures to be taken if irregularity was being caused on any part of a route. A limited type of control is being carried out to-day by observers posted at strategic points on the periphery of the central congested area, who log each bus as it passes the point and advise a central control office by telephone of any irregularity greater than some predetermined interval.

It is unnecessary to enumerate the difficulties facing the bus operating staff to-day, but traffic delays can be divided into two main categories—normal congestion, which results in buses becoming bunched together, and major traffic jams, which cause a complete cessation of movement over considerable periods. When buses have become bunched, the leading bus is usually carrying more than its quota of passengers and is tending to take longer to unload and load at a bus stop; measures can be taken to pass one of the emptier buses from the rear to the front, the bus times then being more nearly equalized. In a more acute degree of congestion, buses about to enter the congested area may be turned back to fill the gaps caused by the normal service being held up in the opposite direction.

Without an overall picture of the location of the buses on a route it is difficult to prescribe what remedial measures are suitable; moreover, such measures may have serious repercussions, since bus crews must be relieved at stated periods, either for meals or at the end of their normal turn of duty, and it is essential that the crews running these reliefs correspond to the garage at which the bus ends its schedule.

There are two other points in the operation of buses which require a constant check. The first is the practice of certain drivers giving their conductors a light run. If a driver on a normal 2 min headway speeds up in the early part of his run, so that he is $\frac{1}{2}$ min before schedule, the bus will average only three-quarters of its normal load and the conductor's work is eased. The following bus, if keeping schedule, will collect some 25% more than its normal quota, and since the time at each stop will be greater, it will become more and more delayed and overloaded. The second point is the passenger who complains that he had an excessive wait for his normal bus on a wet evening. It is most desirable that facts regarding the allegations are available before action is taken.

All the above factors in bus operation can be dealt with if some form of instrument can be produced to signal and record at a central point the identity of each bus as it passes various points not too far apart along the route. This was the problem put to the authors in the summer of 1954, and the outcome is the equipment described in the paper. This has been called the bus electronic scanning indicator (b.e.s.i.).

(2) BASIC REQUIREMENTS

The problem for London Transport is to trace some 7 000 buses operating on 500 routes. An early estimate was that, overall, there would be some 200 scanning points necessitating at least 400 scanning units. The fact that there are 7 000 buses makes it clear that the equipment on the bus must be as inexpensive as possible to make a system attractive economically. It should be mentioned here that any form of oral reporting would be unacceptable, because even if only 1% of the bus crews wished to report at any instant, 70 controllers at the receiving end would be required to accept the messages. Furthermore, owing to the criss-crossing nature of London's bus routes, it would be almost impossible to arrange for groups of routes to report to a single individual, thereby necessitating group controllers to take over from the message-receiving staff. A workable system therefore needs a permanent but up-to-the-minute picture of the traffic conditions, suitably grouped for easy recognition.

With these factors in mind it was finally decided that an optical system would best meet the London Transport needs, but the major difficulties to be solved were, first, to ensure that only the correct light entering the instrument operated the device, and second, that extraneous reflections from other bright surfaces were rejected. It is proposed to base the description which follows on the pilot system which was put into operation initially, since this equipment is easier to follow and is of simpler and cheaper construction than that which has now been developed from it. The reason for the further development and particulars of its construction are given later.

Mr. Pick and Mr. Readman are with the London Transport Executive.

(3) DESCRIPTION OF THE APPARATUS

(3.1) General

There are five basic types of apparatus forming the system, namely the bus identification plates mounted on each bus, the scanning unit mounted near the street kerb at each check point, the information forming and sending unit mounted near to the scanning unit, the receiving unit and the bus-route display panel. The two last items are housed at the control centre, and are connected to the kerb-side apparatus by telephone cable.

The system consists of a combination of optical and electrical techniques and operates as follows (see Fig. 1).

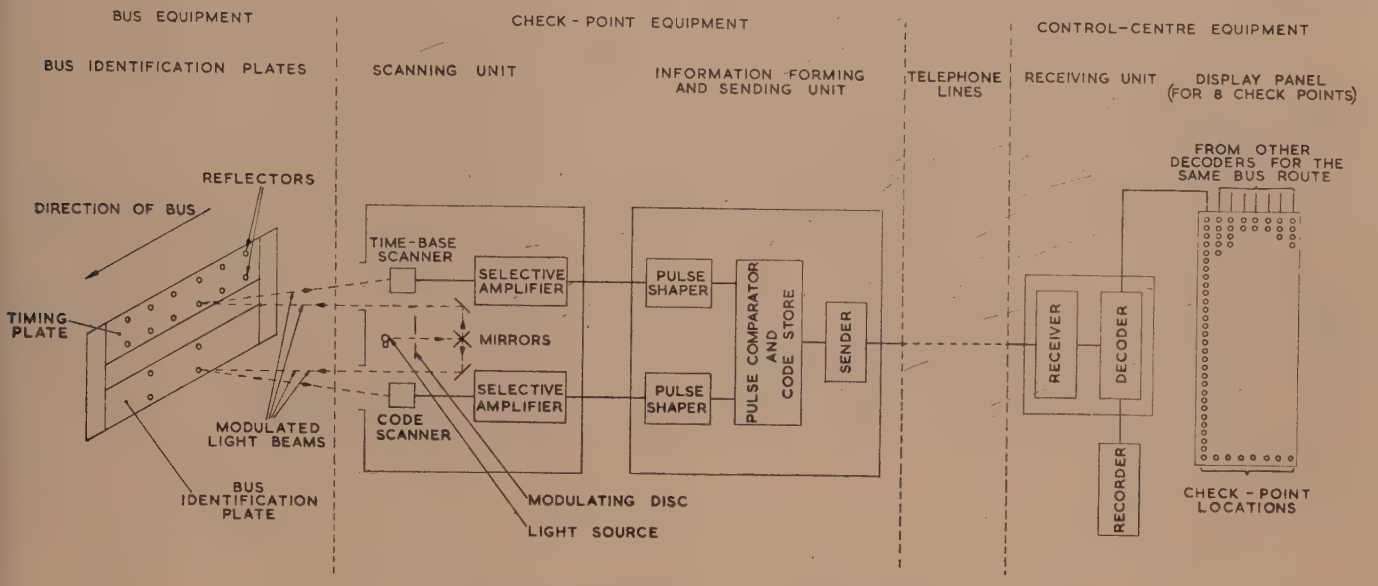


Fig. 1.—Block diagram of double-beam system.

A light source is mounted in the scanning unit, and the beam, after passing through slots in a rotating disc to give it a distinctive characteristic, is divided by a system of mirrors into two nearly parallel beams and projected on to the bus identification plates, consisting of an upper plate with a number of equally spaced reflectors which form the time-base, and a lower identification plate in which the number and position of the reflectors are so arranged as to represent, in binary form, the running number of the bus.

The reflected light is projected back from the upper and lower plates into the scanner through separate windows on to the time-base and code scanners, respectively, where photo-electric cells convert the light into electrical energy. The received light has already been modulated by its passage through the slots of the rotating disc, and the electrical signals will accordingly vary at a known frequency, thus enabling the signals to be filtered by selective amplifiers. Any current flowing in the system due to light from other sources will therefore be eliminated. These signals are passed to the information forming and sending unit, where the pulses are shaped to a suitable waveform for operating valves or relays; and by comparing the signals from the timing and code plates, the binary number of the bus can be built up and passed to a code store, where the signals are either stored for later transmission or are routed to the sender for coding and transmission over a pair of ordinary telephone lines.

At the control centre the code pulses are received, decoded and passed to a display panel, where an indication is given of the running number of each bus as it passes the scanning point. The indication is cancelled when the bus reaches the next scanning point on the route, so that each section of the display panel will give the numbers of all buses between any two scanning points. Provision is also made for a permanent record to be made on a recorder.

(3.2) Bus Identification Plates

The main factor governing the design of the bus equipment is the requirement for a robust and relatively small identity code plate that is readily transferable, and the type used in the pilot scheme consists of two plates with double rows of dome-headed cast-glass reflectors of the type used in road sign cat's-eyes. The $\frac{5}{8}$ in-diameter reflectors are moulded with resin in sheet-metal shells, each row having a 1 in vertical spacing between centres and a 2 in horizontal spacing between the pairs so formed. (Pairing is adopted merely to increase the effective vertical length of each reflector.) The two plates so formed are arranged to slip into guide bars attached to the front near-side of the bus above the driving cab, this being above the level of interference from pedestrians and most other obstructions such as private cars. The plates are held at 10 in vertical centre spacing in the guide bars. The upper plate is the timing reference for the lower plate and has a complete set of reflector pairs. The lower plate is fitted with pairs of reflectors or blanks as are required to form the running number of the bus in binary form. It will be appreciated that, since the bus speed is variable over a wide range, fixed time intervals alone cannot be used as a means of measuring binary figure values. Therefore, an accurately aligned reference between the timing and code plate is an essential feature of the scanning system.

The physical dimensions of the reflectors and their spacing is the determining factor in the design, not only of the arrangement of the scanner light source and optical systems, but also of the type of pulse-handling components used in the circuits immediately following the scanner. A further requirement is the need to ensure satisfactory operation in varying degrees of light from bright sunlight to complete darkness.

A total of 30 buses are needed to operate the No. 74 route, and for pilot trials, therefore, a binary system of five figures covered this requirement. A sixth reflector was added to the timing plate, however, to initiate the transmission of identity information to the central control. The overall space occupied on the side of the bus for this 6-figure code was 14 in × 14 in.

If a larger number of buses on any route must be catered for, this can be achieved by means of additional figures or sets of reflectors in the horizontal line. Furthermore, only the bus identity numbers were incorporated in the pilot system, but if a scanner were required to deal with more than a single route, it would be necessary to add further sets of reflectors to cover the route number in coded form.

The individual reflector buttons have a conical acceptance angle of about $\pm 20°$, which allows ample margin for the tilting of the plates due to road camber, for vehicle loading and for any reasonable degree of variation of angularity of the bus in relation to the kerb line as it passes the scanner.

(3.3) Scanning Unit

(3.3.1) General.

The scanning unit equipment must be erected at the same height as the bus plate and as near as possible to the roadway, either on a post at the pavement edge or, where the pavement is narrow, on a convenient point on the building line. The fixings are permanently mounted and the equipment case is made interchangeable, so that complete units can be replaced for overhaul or adjustment at a central workshop.

The case is completely watertight, with a small amount of ventilation to prevent overheating by the electrical equipment mounted inside. Facilities are provided for the replacement of the light source without disturbance to the rest of the equipment.

In the later designs this unit is 22 in high, 5 in wide and 11 in deep, and weighs about 45 lb.

(3.3.2) Scanner Design.

The light source (Fig. 1) of the scanning unit is a 12-volt 36-watt commercial type of car head-lamp with a small-diameter axial filament. A lens combination, close to the lamp, directs a half-size image of the lamp filament on to the periphery of a disc in which are cut 100 radial slots. The disc is rotated by a small motor at 3 000 r.p.m., and interrupts the beam of light through the slots 5 000 times a second. Two plane mirrors, $5\frac{1}{2}$ in from the modulating disc and inclined at 90° to each other, accept the modulated beam and reflect it in opposite directions to two further plane mirrors, which can be adjusted in angularity to give two beams which emerge from the scanner ports to diverge to about 10 in centre spacing at the average scanning distance of 12 ft.

At 12 ft from the scanner the two images of the filament are approximately 6 in high and $\frac{1}{2}$ in wide. The scanner is adjusted so that one of the beams is centred on the code plate and the other on the timing plate as the bus runs by at the average distance and under average loading conditions. The total change in the vertical direction between an empty bus and one fully laden is about 4 in, so the depth of beam at the reflector plates is ample to cover variations in plate height due to loading as well as variations in height brought about by divergencies of the buses up and down the camber of the road. A wider beam image would be a disadvantage in that the on/off ratio of reflection from the reflector plates, at present nearly unity, would deteriorate, with resulting difficulties in pulse counting.

Each of the two beams reflected back into the scanner is picked up by a lens of 4 in focal length, the two lenses being

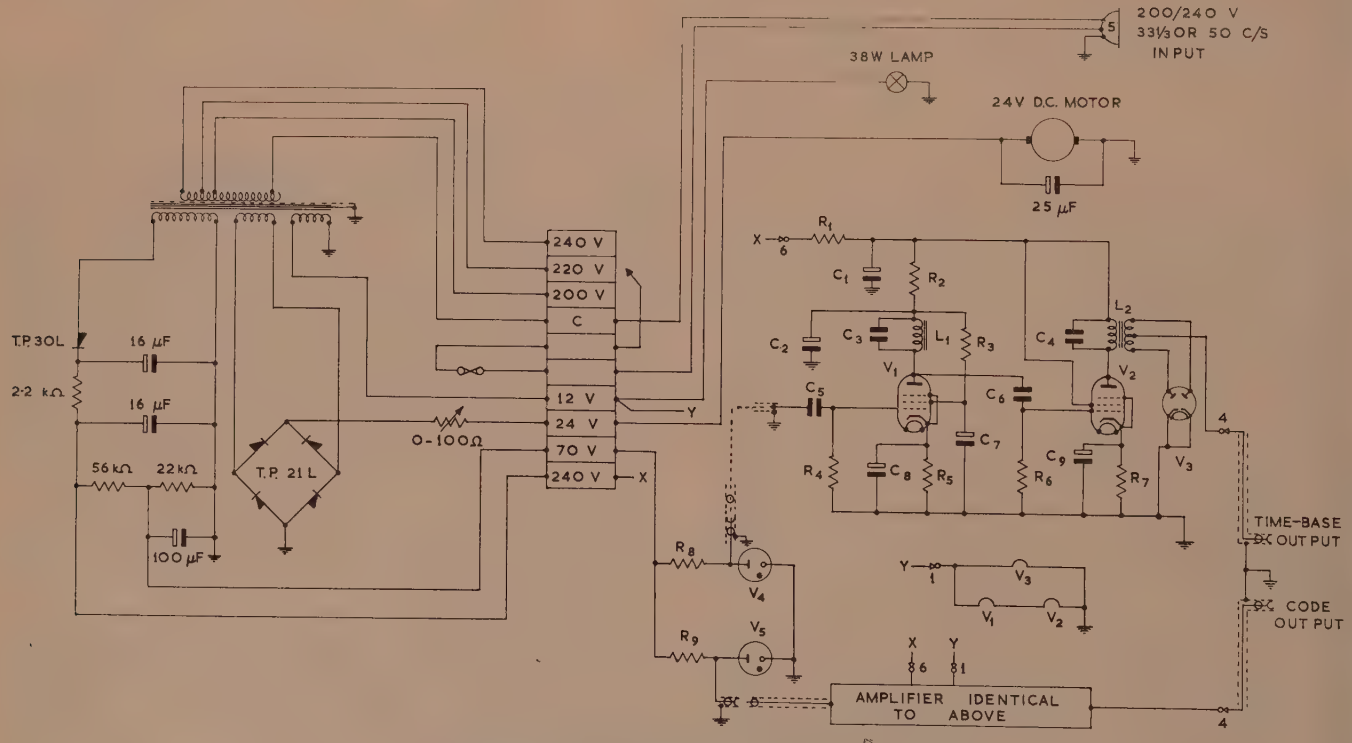


Fig. 2.—Scanner power supplies and selective amplifiers.

R_1	1 kΩ	C_1	16 μF	V_1, V_2	EF91
R_2	22 kΩ	C_2, C_7	8 μF	V_3	12H6
R_3	8·2 kΩ	C_3, C_4	Padder	V_4	90AG
R_4	2·2 MΩ	C_5	0·002 μF	V_5	90CG
R_5, R_7	220 Ω	C_6	0·02 μF		
R_6, R_8, R_9	1·2 MΩ	C_8	50 μF		
		C_9	50 μF		

L_1: 1 700 turns of 38 s.w.g. enamelled copper.
L_2: Primary—980 turns of 34 s.w.g. enamelled copper.
Secondary—2 windings each of 1 250 turns of 38 s.w.g. enamelled copper.

0 in apart. A complete picture of the associated passing reflector plate is thus accepted from each bus, and is, in turn, directed n to two masking plates placed at the focal points of the lenses. The masking plates have central apertures slightly larger than he image of one reflector-plate image, and the plates are so djusted that only the time-reference reflections are accepted y one aperture and the code reflectors by the other. Photo-ensitive transistors behind the masking plates convert the light eams into voltage pulses, which are still modulated at 5 kc/s; he subsequent electrical circuits are shown in Fig. 2.

The complete lens-transistor units are assembled on Paxolin ubes which are mounted within the scanner so that they are djustable in all planes. In practice they are set so that the rincipal axes of the two receiving lenses are parallel and at the ame spacing as the reflector plates on the buses. Code and eference pulse trains from the transistor circuits then drive -stage anode-tuned hard-valve amplifiers, which further amplify he modulated signals and effectively reject the effect on the ransistors of sources of light external to the scanner. Each of he amplified signals, which are approximately sinusoidal, then asses to full-wave valve rectifiers, which remove the 5 kc/s nodulation and provide an output from the scanner of two rains of d.c. pulses some 100 volts in amplitude, and of negative olarity with respect to the scanner casing and earth. The wo outputs from the rectifier circuits are then fed from the canning unit to the information forming and sending unit. The canner casing accommodates, not only the apparatus described bove, but also power supplies for the modulating-disc motor, he light source, transistors and tuned amplifiers.

The width of beam falling on to the identity plates is, of course, ependent on the distance between the bus and the scanner, and at 12 ft the optimum safe working ratio of approximately equal on and off signals is received back. With bus speeds of up to 40 m.p.h. and a 2 in spacing between reflectors, these on and off pulses will be received at the rate of 350 a second, which is well within the range at which the subsequent valves can operate.

(3.4) Information Forming and Sending Unit

(3.4.1) General.

The information forming and sending unit, which accommodates the pulse shaper, the pulse comparator and code store, and the sender, is enclosed in a box measuring 36 in × 8 in × 8 in, and is either made watertight and suitable for pole mounting with the scanner or is placed in a street-type pillar adjacent to the scanner. The pulse shaper modifies the identity code and timing reference pulses to make them suitable for operating the valves or relays at a later stage. The comparator compares the identity code pulses with the timing reference pulses, and the correct binary number is set up in the code store for sufficient time for the code to be transmitted by the sender in binary form as a train of pulses over a telephone line to the receiving apparatus at the central control point.

(3.4.2) Pulse Shaper.

The pulse trains from the amplifiers are received as direct current by the pulse shapers, which convert them to square-wave pulses of constant amplitude for driving the subsequent cold-cathode-valve pulse comparator.

Shaping and limiting is effected by interconnecting a double triode with a voltage reference valve as shown in Fig. 3. Negative pulses of as low an amplitude of 15 volts will drive the shaper and provide the requisite type of output pulse. It is interesting

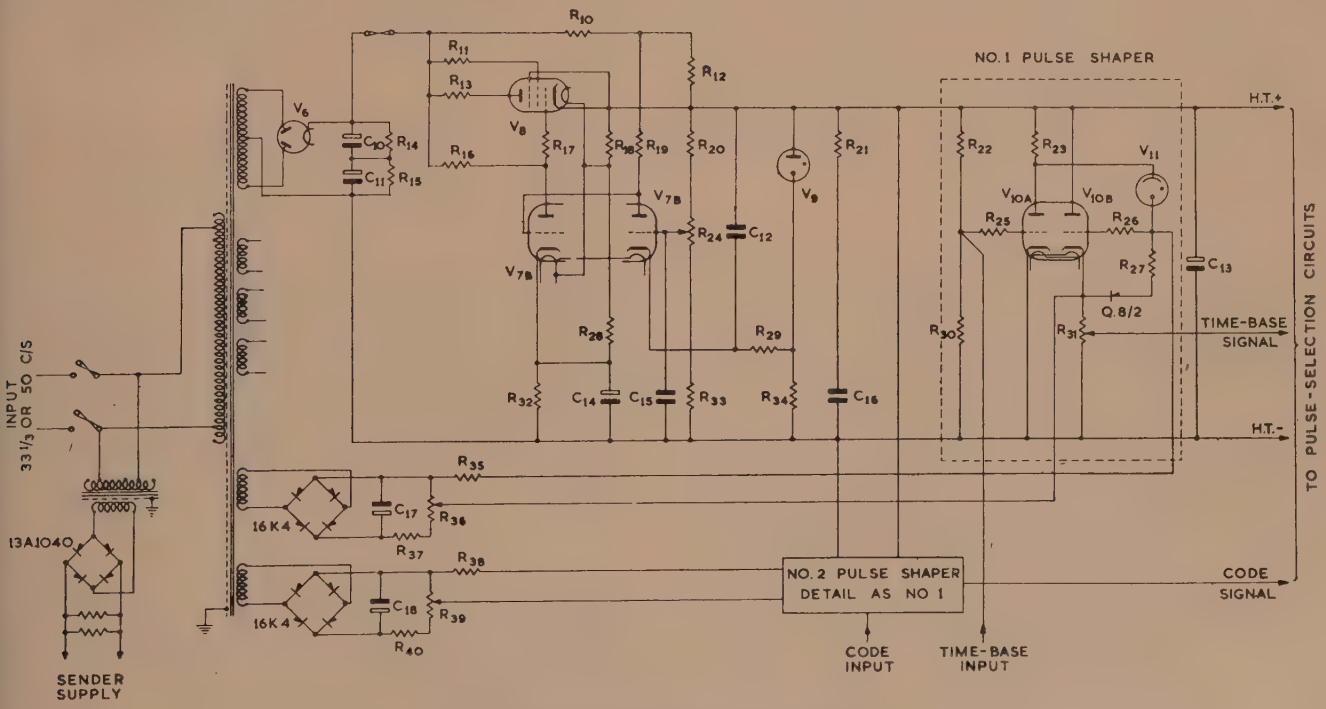


Fig. 3.—Pulse shapers and power supplies.

R_{10}	150 kΩ	R_{21}	0·5 Ω	C_{10}, C_{11}, C_{13}	32 μF	V_6	5U4G
R_{11}, R_{13}	100 Ω	R_{22}	2·2 MΩ	C_{12}	2·0 μF	V_7	12AX
R_{12}	1·8 kΩ	R_{23}	47 kΩ	C_{14}	2·0 μF	V_8	6CH6
R_{14}, R_{15}	1 MΩ	R_{24}	15 kΩ	C_{15}	0·5 μF	V_9	QS1200
R_{16}	1·2 MΩ	R_{25}, R_{26}	270 kΩ	C_{16}	4·0 μF	V_{10}	6SN7
R_{17}	10 kΩ	R_{27}, R_{28}	22 kΩ	C_{17}, C_{18}	2·0 μF	V_{11}	XC18
R_{18}	12 kΩ	R_{29}	1 kΩ				
R_{19}	390 kΩ	R_{30}, R_{35}, R_{38}	330 kΩ				
R_{20}	120 kΩ	R_{31}, R_{36}, R_{39}	20 kΩ				

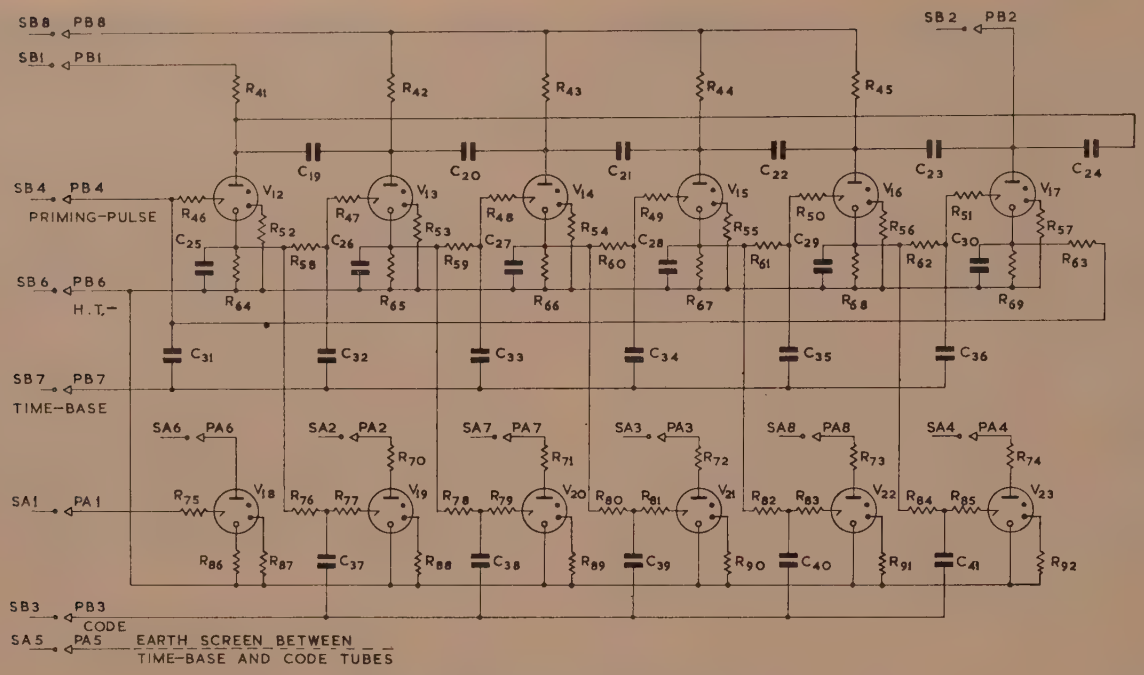


Fig. 4.—Pulse comparator and code store.

R_{41}	8·2 kΩ	R_{64}–R_{69}	33 kΩ	R_{77}, R_{79}, R_{81}		C_{19}–C_{30}	0·2 μF
R_{42}–R_{45}	15 kΩ	R_{70}–R_{74}	22 kΩ	R_{83}, R_{85}	270 kΩ	C_{31}–C_{41}	2 000 μμF
R_{46}–R_{51}	270 kΩ	R_{75}	3·9 kΩ	R_{86}	39 kΩ	V_{12}–V_{23}	Z700U
R_{52}–R_{57}	10 MΩ	R_{76}, R_{78}, R_{80}		R_{87}–R_{92}	10 MΩ		
R_{58}–R_{63}	1 MΩ	R_{82}, R_{84}	1 MΩ				

to note that, under favourable conditions, the scanner will produce a pulse-voltage output more than six times that required for the input to the shapers. A considerable amount of deterioration, both in the scanner and in the reflectors on the bus, can thus be accommodated before the equipment fails to register identity codes correctly, and it is hoped that in conditions of bad visibility, such as fog, the equipment will operate so long as the buses themselves continue to run.

(3.4.3) **Pulse Comparator and Code Store.**

The circuits for selection and storage are shown in Fig. 4, and those for sending in Fig. 5. It will be seen that selection is carried out by two rows of cold-cathode trigger valves. The square-wave timing reference pulses supply the common trigger line of a chain of six cold-cathode valves, and the first valve in the chain is prepared by a preceding priming circuit which raises the trigger potential amost to striking point.

The first timing pulse triggers the first valve, which in turn extinguishes the primer by reason of the fall in anode potential caused by the intentional introduction of poor regulation characteristics in the common anode circuit. Conduction through the first valve also prepares the trigger of the second and at the same time biases the code valves mentioned later. Repetition of this sequence produces a transfer from valve to valve in step with the incoming timing reference pulses, the last pulse operating a relay in the anode circuit of the sixth valve in the chain, which, in turn, initiates the line sending circuits.

Similarly, the modified code pulses are supplied to the common trigger line of a further chain of five cold-cathode valves which are in turn supplied with a biasing pulse from the associated timing reference valve. The code valves are therefore made conducting in the correct order of the incoming code train pulses and either fire or not according to the figure or cipher of the binary number. The valves which fire actuate 3000-type relays in their anode circuits; thus the code is stored until the sending sequence commences. When the complete signal has been stored, the whole receiver circuit is cleared down and the first valve is primed in readiness for the next train of pulses.

(3.4.4) **Sender.**

In the sender, conventional telecontrol techniques are employed using standard 3000-type relays, both in the transmission of the codes to the control equipment and in the reception and decoding which follow. The need for economy in transmission-line conductors and associated equipment leads to the choice of a sequential type of transmission, and because of the relatively long interval between consecutive buses, there is no need for the transmission to be at high speed. The stored binary information is transmitted as a series of positive and negative 50-volt pulses at a constant speed of 10 pulses/sec. The pulsing train consists of six pulses, the first five containing the identity information and the last—which is always a negative with respect to line A (Fig. 5)—being a resetting pulse. Positive is identified with on and negative with off in the transmission of the binary information.

From Fig. 5 it will be seen that the relays TA, TC and TD are used to switch the line at the constant rate required. The polarity impressed on the line will depend upon the position of the contacts of the code storage relays G–L, which are shown in the lower half of the diagram, all in the relay off position. Relays TN–TS are used for moving the line connections along the G–L contacts in synchronism with relay TA.

(3.5) **Connections—Check-Point to Control-Centre Equipment**

As already mentioned, ordinary telephone lines are used for the connection between the kerb-side and the control-centre equipment. Dependent on the storage capacity provided in each code storage unit and the sending speed of the signal, a number of scanning points can be grouped on to a single pair of telephone lines. The line current need never exceed 5 mA o

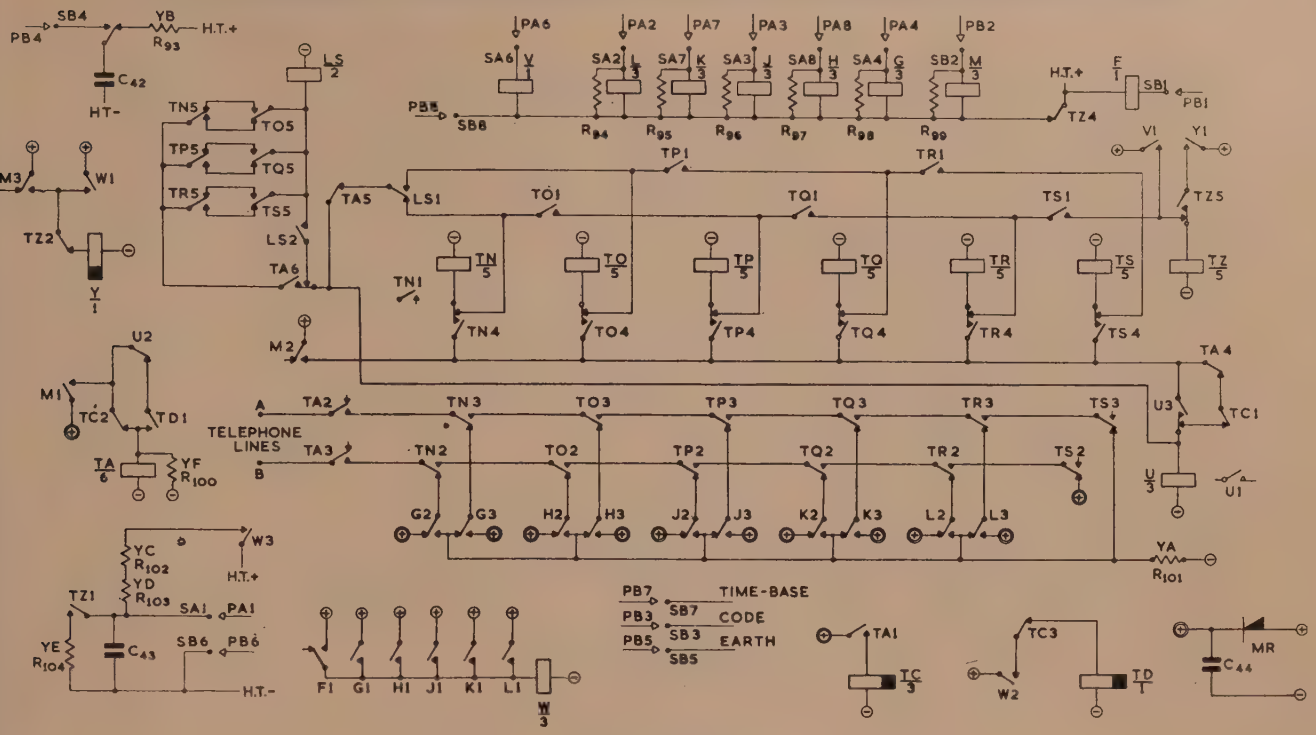


Fig. 5.—Sender.

R_{93}	27 kΩ	C_{42}	0·1 μF
R_{94}–R_{99}	330 kΩ	C_{43}	0·5 + 0·5 μF
R_{100}	750 Ω	C_{44}	16 μF
R_{101}	75 Ω		
R_{102}	1·5 MΩ		
R_{103}	1·8 MΩ	MR—Type H18–16–1	
R_{104}	2·2 MΩ		

Relays

Designation	Coil resistance	Springset combination			
		M	B	C	K
	kilohms				
F	7	—	—	1	—
G, H, J, K, L, M	20	1	—	2	—
LS	2	1	—	1	—
TA	1	1	2	2	1
TC (¼ in h.e. slug)	2	—	3	—	—
TD, Y (¼ in h.e. slug)	0·8	1	—	—	—
TN, TO, TP, TQ, TR, TS	1	1	—	3	1
TZ	1	1	2	1	1
U	2	1	1	—	1
V	20	1	—	—	—
W	2	3	—	—	—

ither positive or negative polarity with reference to one line nd is therefore very much below the maximum direct current vhich may be applied to a Post Office line. The rate of reversals f both current and voltage at 10 per second is also moderate y telephone engineering standards. Furthermore, it is expected hat the background noise associated with the pulse signals will ot seriously interfere with speech over the same line, so that he controller will be able to speak to the roadside staff, or vice ersa, while signalling is in progress.

(3.6) Receiving Unit

3.6.1) **General.**

The information required at the central control point is three-old: first, the attention of the controller must be drawn to any ajor irregularity of the service; secondly, an up-to-minute ecord of the position of each bus operating on the service; irdly, a permanent record of the passage of each individual us and the time at which it passes the check-point.

The first requirement can be achieved once a record of each us passing each of the scanning points is made at the central ontrol, when any major irregularities would be noticed. For stance, the passage of each bus can be made to trigger a timing elay, and if the interval between successive triggerings exceeded given time, a warning signal would be given.

Equipment consisting of display panels has already been developed to satisfy the second requirement. Each panel relates to a particular scanning point, and the panels are grouped together to show complete bus routes; they display the appropriate number of lamps corresponding to the number of buses on the route concerned, and the lamps illuminate numbers corresponding to the identification of the buses. As each bus progresses from one scanning point to the next the lamps light up, signifying the pair of scanners between which the bus is located at any instant of time. The characteristic of the final circuit for illuminating the lamp is so designed that, if two lamps are in circuit at the same time, as when a bus passes from one scanning point to another, the current demand causes the voltage to fall and the previously illuminated lamp is extinguished, leaving the subsequent lamp in charge and only the single up-to-date indication on the appropriate panel.

The third requirement can be met by using a recording tape, or similar device, the tape being punched with the code number of the bus and a further set of punched holes giving the coded time as the signal from each bus passing the scanner is received in the control centre.

The equipment at the central control as used on the pilot scheme comprises a receiver, a decoder, a display panel for each check-point and a recorder.

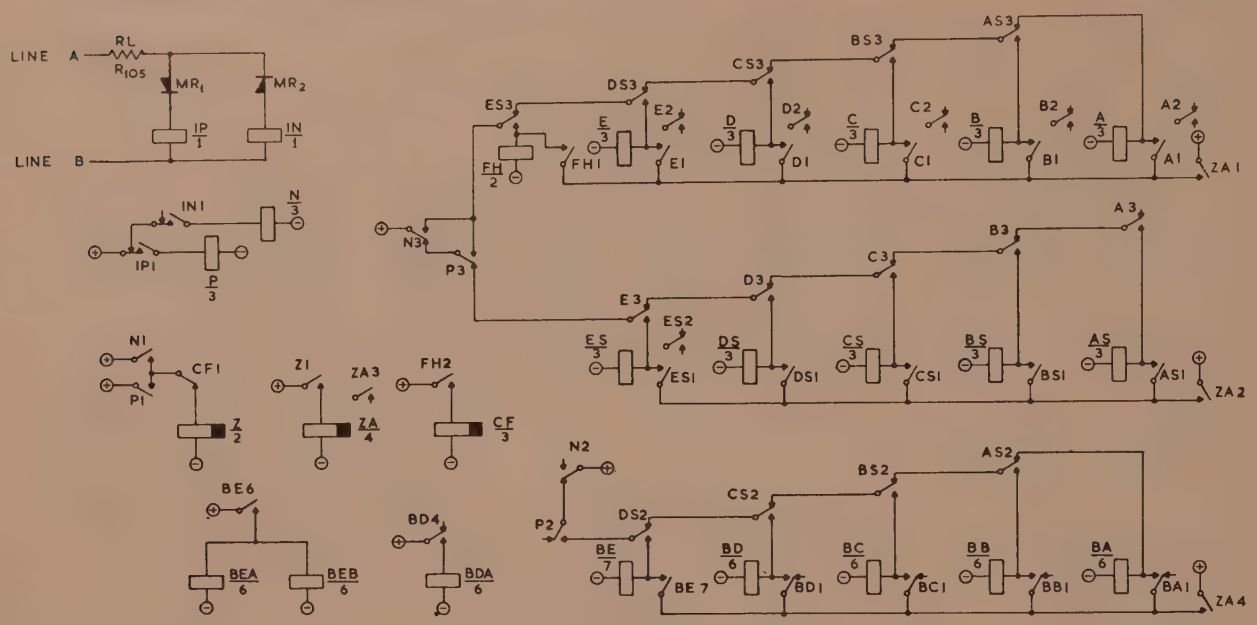


Fig. 6.—Receiver and decoder.

R_{105} 6–16 kΩ MR_1, MR_2—Type B18–5–1 *Relays*

Designation	Coil resistance	Springset combination			
		M	B	C	K
	kilohms				
A, B, C, D, E, AS, BS, CS, DS, ES	1	1	—	2	—
BA, BB, BC, BD, BDA, BEA, BEB	1	—	—	6	—
BE	1	3	—	4	—
CF (1 in h.e. slug)	2	—	1	2	—
FH	2	2	—	—	—
IP, IN	6	—	—	—	1
N, P	1	1	—	—	2
Z (1 in h.e. slug)	2	2	—	—	—
ZA (1½ in h.e. slug)	0·8	4	—	—	—

(3.6.2) **Receiver and Decoder.**

The circuits for the receiver and the decoder are shown in Figs. 6 and 7.

A check-point receiving circuit consists simply of the two rectifier polarized relays (IN and IP in Fig. 6), connected across the incoming line in opposition. Incoming pulses from the sender at the other end of the line will operate one or other of the two line relays in accordance with the polarity make-up of the pulse train. Two further slave relays operated directly from the line-relay contacts are arranged to switch in sequence a chain of time-base relays in such a manner that the binary code train being received on the line can be represented by the on or off switching of a further set of decoding relays.

Decoding is effected by a conventional tree type of interconnection of the decoder relay contacts in which the simple combination of five sets of switch contacts can be arranged to produce a signal in any one of the 31 outlets. The decoder tree-outlet contacts are wired directly to the lamp transfer circuits, as shown in Fig. 7.

(3.7) Bus-Route Display Panel

The complete display panel for the pilot system for the No. 74 route, the circuit for which is shown in Fig. 7, contains eight sections each having 30 miniature fluorescent lamps. Each section is coupled by telephone line to its associated check-point equipment, and the only interconnection between the sections on the panel will be the common power-supply line and the suppl transfer lines.

(3.8) Recorder

The recorder at present in use at the receiving centre is standard type of d.c. voltage recorder driving a paper chart a 3 in/hour. Pen deflection occurs with every incoming identit code, and the magnitude of the deflection is arranged to be function of the bus running number, e.g. a full-scale deflectio represents bus No. 30. This is achieved very simply by makin a set of slave decoder relays (see Fig. 7) switch appropriat resistors into the recorder line.

A minor alteration in the decoder contact wiring would perm direct operation of commercial punched-tape machines.

(4) IMPROVEMENTS AND DEVELOPMENTS

(4.1) General

It is now known that the initial experiments were carried ou in conditions which were more favourable than can be expecte as a general rule. More extensive tests have shown that uneve loading of the bus, or in certain places a bus pulling away fro the kerb, could cause sufficient tilting between the front and th rear of the vehicle to throw the two identification plates out alignment vertically, with a resulting loss of coincidence betwe signals and a failure of correct indication of identity. Moreove the adjustment of the two individual beams of light proved to

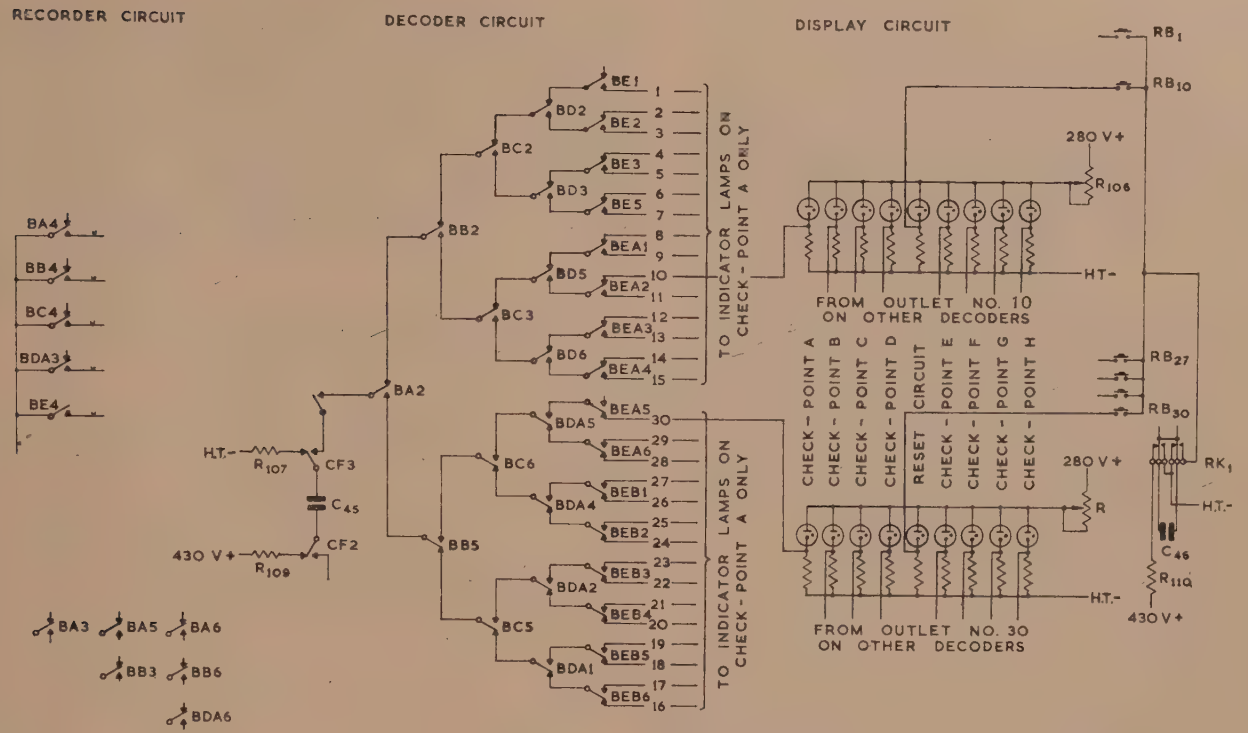


Fig. 7.—Bus-route display panel.

R_{106}, R_{108}, 1 MΩ
R_{107}, R_{109}, 22 kΩ
R_{110}, 4·7 kΩ
Indicator lamp, Type F2
Indicator-lamp cathode resistors, 47 kΩ

difficult matter, and, in turn, this had repercussions on the ethod of mounting the scanner, which was unduly complicated ereby. In order to overcome these difficulties, a modified canner has been developed in which only a single beam of ght and a single reflector plate are required, thus doing away ith the alignment and synchronizing difficulties of the double-eam scanner.

Both the light-transmitting and the light-receiving optical nits have been modified for the new scanner. The transmitting nit retains the motor-driven modulator disc, but the beam-litting mirror system has been omitted and the lens system ow projects a single narrow strip of light directly from the odulator disc to the bus. A cylindrical lens has been added the projector in order to increase the vertical length of the eam.

A single reflector plate on the bus is now used identical in ze and number of reflectors to that used previously for the ming reference. Clear reflectors form the units and red reflec-rs the ciphers of the binary code, and in this way the separate ming and coding plates have been combined. It follows from e adoption of a two-colour reflector plate that the receiving nit in the scanner must be colour selective, and this has been hieved by directing the return beam to the scanner simultane-usly on to two photocells with appropriate colour-discriminating aracteristics. The amplifying and pulse-comparator circuits llowing the photocells are similar to the double-beam scanner rcuits.

(4.2) The Single-Beam Scanner

The arrangement of the optical system of the single-beam anner is shown in Fig. 8 and an isometric view of the unit with the cover removed in Fig. 9. It will be noted that the optical unit, power pack and twin amplifier units have been rearranged to be one above the other, with the optical unit at the lowest level. The three units are attached to the back-plate of the casing, thus allowing for the removal of the main casing shell for adjustment and maintenance.

The only change in lenses in the optical unit is the addition of a cylindrical lens after the modulating disc and the omission of one of the receiving lenses. The cylindrical lens* had been introduced to increase the length/width ratio of the beam projected on to the bus and thus allow more tolerance of bus movement in relationship to the scanner. As mentioned earlier, the beam-splitting mirrors are no longer required, and the projector part of the optical unit has accordingly been reversed and the light source is now in direct line with the code plate on the bus.

The single code plate, measuring 2 in × 14 in, comprises a combination of clear and red reflectors according to the code number required. Back at the scanner, the reflected signals are received in an object lens and are then directed by two plane mirrors, inclined at 150°, simultaneously on to the photocathodes of a type A and a type C photocell. The type A photocell, which is particularly sensitive to daylight radiation with a blue predominance, responds to reflections from the clear reflector but not the red. The type C photocell is most sensitive to incandescent light sources and to near infra-red radiation, and therefore responds to reflections from both clear and red reflectors. It can therefore be seen that, if the binary code make-up on the plate consists of clear reflectors for the units in the code and red reflectors for the ciphers, the type A photo-

* This is actually a lens in the shape of a sector of a cylinder.

cell output will contain only code-unit information and the type C photocell output will give a complete count of all the reflectors in the plate as a timing reference. This arrangement gives adequate separation and obviously complete synchronism between code and timing reference. The outputs from the two photocells, which are similar in character to those in the double-beam scanner, are fed to resonant tuned amplifiers, pulse shapers and a pulse-comparator unit, as already described for the double-beam scanner equipment.

(4.3) Guard Circuits

Little attention has so far been given to protection against the transmission and reception of false information. The electronic timing circuit which guards against undue delay in code transmission (see Fig. 5) could well serve a similar purpose at the reception end of the system.

It is now thought that the code plate should contain preliminary 'clear down' reflection and that conventional mutilated-code safeguards should be incorporated in the transmission circuits.

(4.4) Permanent Coding

The fact that cat's-eyes of two colours can be distinguished makes it possible that a counting system incorporating three colours might be used. This would allow a ternary coding system to be used, which would be attractive when large numbers had to be considered—as would be the case if a system of permanent, rather than transferable, coding was adopted for the 7000 buses operated by London Transport. The possibilities of such a system are now being investigated.

(4.5) Developing and Using the System

A working model of the new indicator system has been built as an aid to the development and testing of alternative forms of terminal indication and recording equipment. This is similar to the full-size equipment, except that modulation of the light beam has not been incorporated, since this is not necessary for the model, which consists essentially of a moving 12 ft track loop equipped with 30 model buses. The buses carry binary-code and time-base plates, and eight miniature double-beam scanners are spaced round the loop. Buses can be stopped and started either automatically or manually from a central control panel. The output from the miniature scanners is taken to a full-size terminal equipment cabinet. The model is ultimately intended for training control personnel in the use of the new indicator system.

(4.6) Other Uses

So far, the development has been confined to application by London Transport. The system would appear to have a wider range of utility, such as recording for sorting purposes any miscellaneous group of moving objects, e.g. the sorting and marshalling of pit tubs at a colliery.

If speeds higher than 40 m.p.h. have to be dealt with, various adjustments can be made to keep the ratio between the modulation and the signal pulses at the preferred level. Larger cat's-eyes and wider spacing between them could be adopted, or probably a more satisfactory method would be to increase the modulation rate of the outgoing light.

There are many ways in which the processing of signals received at the control point can be developed. The most logical first step will be a punched-tape record which could be transferred to punched cards for analysis in some form of computer-type equipment, for the rapid production of mileage statistics, etc.

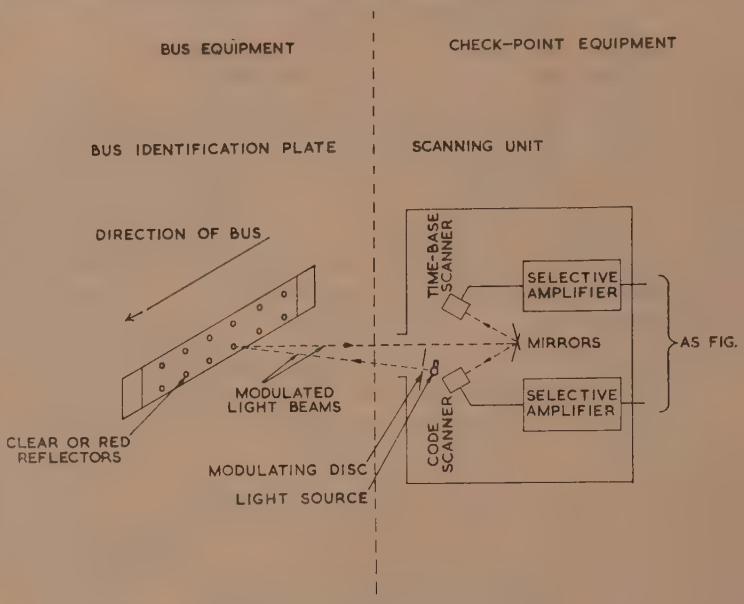


Fig. 8.—Block diagram of single-beam system

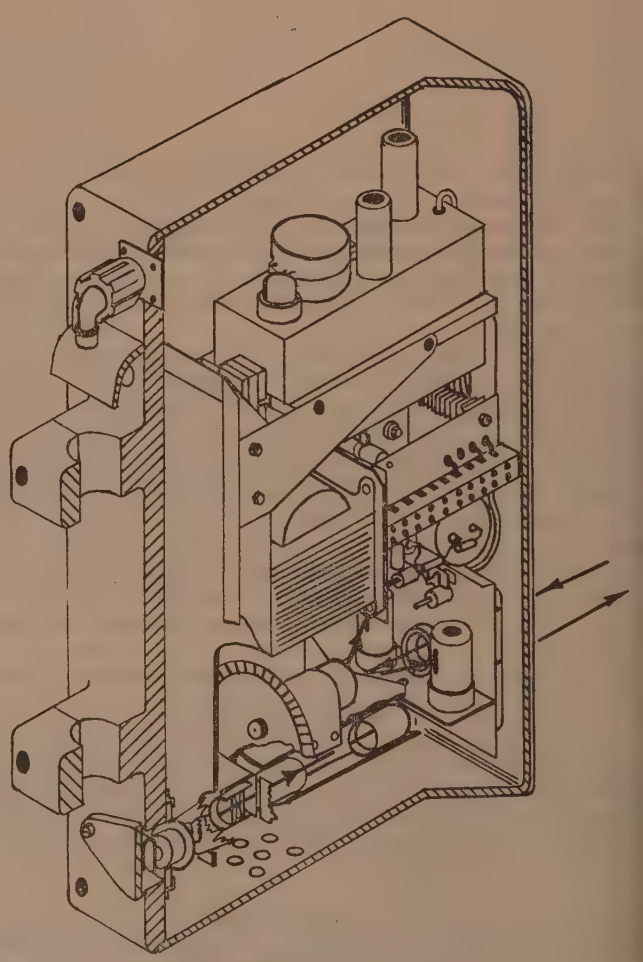


Fig. 9.—Single-beam scanner.

(5) CONCLUSIONS

The problem which has been dealt with by the system described n the paper presented many difficulties, most of which are unique o the system for which it was developed. In this connection it nay be noted that the development of the system for the identi- ication of buses involved variables such as speed of the object, isibility, distance from scanner, changes from the normal plane oth horizontally and vertically, together with the need to reject he effects of reflections from sources other than from the code lates—a combination of complications which it is highly mprobable would be encountered in the application of the system n any other sphere.

The electronic bus scanning device is covered by provisional patent application No. 33758, 1957.

(6) ACKNOWLEDGMENTS

The authors wish to express their thanks to the London Transport Executive for permission to publish the paper and to members of the staff of London Transport concerned with the development and production of the equipment for the bus electronic scanning indicator, as well as to many manufacturers who have given great assistance in advising and recommending the best units to build up the assembly.

DISCUSSION BEFORE THE INSTITUTION, 6TH NOVEMBER, AND THE MERSEY AND NORTH WALES CENTRE AT CHESTER, 24TH NOVEMBER, 1958

Mr. C. C. Inglis: The technique employed, and in particular he use of the 2-colour system—which, in effect, introduces a hird dimension—is very ingenious. In this kind of develop- nent there is a general question which is important, and I should ike to ask the authors how the problem as a whole was pproached, because there may be a kind of vicious circle: the ngineer does not quite know what the user wants, and the user oes not quite know what the engineer can provide. Did the uthors or the operators break this circle? Did the operators ay, 'We must have something to tell us where the buses are', r did the authors say, 'We can provide you with something to ell you where your buses are'?

I should like to mention a very similar idea, on a much larger cale, to record the position of every wagon on British Railways, nd the time. This started by the engineers saying that they ere going to develop a technique of solving this problem, ithout having the assurance that the operating and commercial ides wanted it. The operators said, 'Show us that it can be one, and we will then think about how to use it and whether r not we really want it'.

The authors have shown that the problem could be solved. Whether or not their system is wide enough for our million agons, covering the whole country, I do not know. There re other methods of solving the technical problem. But then omes the question, 'When we have the information, what shall e do with it?' That is a vast problem and affects almost the hole railway system including communications, and I should ke to know whether the authors use the Post Office lines or ne London Transport lines for communication.

Referring again to British Railways, let us assume that, of ne million railway wagons in this country, half rank for this ind of treatment; and let us suppose that we want to record neir position three times a day. This means 1 500 000 signals oming in every 24 hours, which have to be stored and processed. s yet we have not made an assessment of the size of the machines ecessary to do this, and I shall not make a guess at this stage. here is no doubt, however, that the techniques the authors have olved and the experience they have gained will be of very great alue to us in the future, provided that we are sure that we want equip 500 000 wagons with a system of this kind.

Mr. A. R. Enshaw: On the face of it, the paper would appear give a comparatively simple answer to a not-so-simple roblem, but I have doubts whether this simple light source will ope with certain weather hazards. The penetrating power of ght is the essential factor, and it has to be adequate for all easonable climatic conditions. Is it not possible to apply the ience of physics to an even greater extent than the authors ave done? It may be that what they have done is adequate, ut I am sure that they would not wish it to be considered a panacea for the general problem of recognizing and recording moving vehicles by electronic means.

My company has been employed on a somewhat similar task of recognizing and recording information about vehicles which may be passing at speeds up to 80 m.p.h. The equipment has to operate in conditions of driving rain, sleet or dense fog, and has to withstand heavy-handed workmen. Furthermore, the information which is collected from these passing vehicles must be stored for long periods—perhaps years. Facilities must be provided giving automatic access by inquirers to all the stored information. Moreover, where the authors' electronic indicator has the problem of dealing with 7 000 buses, my problem is the recognition and recording of up to half-a-million vehicles.

I have elected to make this comparison because the bus indicator might be considered to be a fine-weather device, whereas the scheme with which I am concerned has to operate in all conditions. It is a similar problem, but a much more complex one, although the field trials have so far been satis- factory. It is now a question of whether the application will justify the cost. The authors have shown that they have achieved their end, and I wish them every success.

Mr. B. H. Harbour: I am a member of the London Transport Executive and the leader of the operating departments which the authors of the paper are serving in the development of this device. The need for improved communications in the control of London Transport road services has long been recognized, and this need initiated and stimulated the development of the equipment described.

When one thinks of the mass of telecommunication and signalling on the Underground system, and compares it with what we have for the road services, the need is obvious. We now control our buses by what may be called manual methods and the use of a congested telephone system. We in London Transport feel that this development will improve communica- tion and be of major advantage in the control of London's buses. In saying this I must add a warning, on which a previous speaker has touched. The requirement is not simply for more information: there must also be effective utilization of the information. Control techniques are developing, and this development will make rapid strides once equipment of the type described is at our disposal.

There are two points which I would put to the authors. First, we must look for a high degree of reliability in this equipment, which will be required to function efficiently under all weather conditions, including fog, ice, snow and rain. If the bus indicator is to be used for a more centralized control, involving fewer men, any failure—as with a signalling failure on the railways—would be serious. Secondly, the authors make no reference to costs. There are limits to the extent to which under-

takings can go in the development of new apparatus, and the cost factor is of importance. Perhaps they will also touch on this aspect in their reply.

Mr. D. F. Nettell: It would seem that the authors' equipment, containing as it does hard valves, gas tubes, relays, transistors, photo-electric cells, chopper amplifiers and electric motors, should provide interesting evidence in the perennial controversy of relays versus valves. Can the authors, from their experience of this equipment, give any figures for the comparative reliability of the various components, including resistors and capacitors?

It would seem that they have made the best possible use of the various components in the equipment. They have used hard valves where information rates are high, gas tubes and relays for comparison and storage, and relays for the complex logical functions of control and to transmit information at comparatively low speeds. It would be interesting to know, however, whether any consideration was given to designing equipment with one type of component throughout, e.g. transistors.

Secondly, it would seem that, since the train of events which culminates in the transmission of information to the central point and its decoding and display there is set off by the passage of the bus, and—with due respect to London Transport—in this context the passage of the bus is a purely random event, one telephone circuit has had to be used to connect each roadside station to the central point, and one set of receiving and decoding equipment employed there for each circuit. This may be satisfactory for an experimental installation, but, bearing in mind what previous speakers have said, for a comprehensive installation I should be interested to know whether consideration has been given to multiplexing or time-sharing the receiving equipment and to the possible use of party lines in the telephone circuits.

I can appreciate the fact that the controller at the central point, with an illuminated display, has a very good idea of what is happening on route 74, and can see bunching at some points and gaps at others, but how does he use this information to remedy the situation? It seems to me that the controller must be able to communicate with each individual driver; there must be some feedback of information. How is this to be done?

Mr. P. F. Cook: One of the most interesting things about this system is the extreme simplicity of the installation on the mobile unit, the bus. Inductive signalling would appear at first sight to be the more obvious solution, and in these days of semiconductors one would think that it might be possible to carry a fairly simple and cheap transistor oscillator-generator on the vehicle. I imagine that this must have been fully considered, and it would be interesting to know the reasons for rejection of such a system.

Information on the reliability of the system would be interesting. Is the probability of a miss known with accuracy? What is the range of the system and what are the factors which limit this? In this connection, one would imagine that the signal/noise ratio could be substantially improved by using a phase-conscious demodulator in place of the simple rectifier. This would be easy to achieve if a synchronous motor were used to drive the scanning disc—an arrangement which would probably be more reliable than the present d.c. motor.

As regards reliability, it is noted that the system depends on tungsten-filament lamps (which usually have a life of only 1 000 hours). Are any arrangements made to improve this? It is also noted that the valve types specified are not those known to be the most reliable.

Finally, how does the system react to heavy snow conditions? There will be the interference of large snowflakes in the 2-way beam system and also the danger of snow collecting on the cat's eyes themselves. In normal use, cat's eyes have what may be described as a built-in wiper, and it would be interesting to know whether those on the bus can, in fact, be kept sufficiently clean.

Mr. G. Bishop: In my early days I was in the telephone world, and I recall that the number of first selectors provided was based on the probability of the number of lost calls per busy hour which could be allowed. It seems to me that the same thing will apply to this apparatus; it will be a question of the number of lost codes per rush hour which can be allowed. A furniture van will go by and the first scanning post will miss the bus; a pigeon will fly by at the critical moment and the second scanning post will miss the bus. The chances of catching the bus at the third scanning post will be good. It means that the amount of apparatus to be provided in the way of scanning units will have to be based on the probability of the number of lost buses which can be allowed per rush hour.

Mr. C. J. O. Garrard: The problem of dealing with the huge amount of information which this system will collect if and when it is installed on all the bus routes in London is likely to be difficult. Similar problems have arisen in my own experience in trying to apply computers to materials control in a large works. One sees that certain information requires recording and processing; with expert help it is not hard to devise suitable mechanism. The trouble starts when one is confronted by the resulting flood of figures. In most commercial systems one is not concerned when everything goes well; one asks to be warned when something goes wrong. One therefore records only the departures from the norm. Has this thought crossed the authors' minds? Could they operate their scheme by setting up in the central office the correct route timetable and recording only the departures from it?

My second point concerns the storage of data, and I address myself not so much to the authors as to computer designers. The problem of data storage has been solved at both ends, but not in the middle. We have available very quick stores for moderate amounts of information. These have access times of the order of microseconds, but will accept only amounts of information which are small compared with those of which the authors and Mr. Inglis are thinking. On the other hand, stores for very large amounts of information, such as magnetic tapes, have access times which run into at least tens of minutes.

What we want for many problems, and perhaps for the one with which we are dealing here, is a store for the order of several million bits with an access time, for both recording and reading, of the order of seconds. So far as I know, this is a problem which has not yet been solved; it deserves attention.

Mr. P. P. Dwivedi: Is there any indication when the apparatus fails to work correctly? This indication should be a great help in correct interpretation of the codes at the receiving end. Would the authors think it desirable to insert a circuit to ensure the reception of the desired signals only, before the formation and the transmission of the intelligence?

With regard to the motor, if the scheme is applied to more buses, or if a greater amount of the information has to be fed simultaneously, how can it be dealt with in the present design? Has simultaneous reception been tried, and, if so, with what degree of reliability and accuracy?

Mr. R. D. Haigh (*at Chester*): No rotating parts are used in this equipment except the scanning motor, and it is rather surprising that a high-speed d.c. motor should be chosen for this purpose. Surely maintenance costs would be considerably reduced if a lower-speed a.c. motor were used?

It appears to me that there are at least the following two possible sources of error in the information obtained from the equipment, even though it is itself functioning quite correctly:

(*a*) Buses overtaking each other or a high vehicle at the moment of passing the scanning point.
(*b*) A bus coming to rest while in process of being scanned, moving a few inches backwards—a possibility even with the best of drivers—and then finally moving forwards again.

The latter is probably sufficiently unlikely to be discounted, but the former seems to me to be a very real source of error if the system is in use on a section of road carrying several bus routes and many commercial vehicles of sufficient height to interrupt the scanner beam. The danger may be reduced by the fact that the scanning distance is limited, but it is a factor which cannot be overlooked.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. T. S. Pick and **A. Readman** (*in reply*): In reply to Mr. Inglis, the problem of bus recognition was first posed by the operators, who stated their requirements and then co-operated with the engineers in developing the system. The question of communication channels has been met in the case of the Route 74 experiment by using private lines, but it is likely that some Post Office lines may be used as the system expands.

With regard to the wagon problems on British Railways, we consider that, in so far as transfer and transmission of identity information are concerned, the bus system could well be widened to meet the larger number of vehicles.

Storage and utilization of information on the scale indicated by Mr. Inglis is outside the scope of our system, which is primarily concerned with the presentation of information for immediate use, although some provision has been made for information processing on standard punched-tape machines.

Mr. Enshaw raises the question of the scanner light-source and its penetrating power. We agree that there could be a failure in extremely bad weather conditions, but tests so far this winter have indicated only about a 10% reduction in scanner output due to fog. Figures of the effects of heavy rain and snow have not yet been taken.

Mr. Harbour emphasizes the need for improved signalling facilities for road-service vehicles, and the system described is designed to meet that need with minimum cost. It has been estimated that the capital cost of equipment in the final arrangement should be considerably lower than the annual cost of an equivalent manual control.

It is a little early in the trials to assess, for Mr. Nettell, the comparative reliability of components in the equipment. In the six months that the equipment has been running, certain of the valves have given trouble and have been changed for a different type. Little trouble has been experienced with other components.

Solid-state switches may ultimately replace the valves and relays, but the comparative cost of such devices is high at present and certain of their advantages, e.g. compactness and low power consumption, are not of great benefit in this application.

The question of the efficient use of equipment and telephone lines has been considered, and we feel that multiplexing will be an essential feature of parts of the final communication system. The receiving equipment will most likely be a combination of route-sorting units with individual route receivers and display panels.

There will be no controller-to-driver communication. Information from the controller will be telephoned to one of a number of inspectors, located at strategic points on the route, who will initiate any required service changes.

In answer to Mr. Cook, inductive signalling has been tried and rejected on the grounds of cost and difficulty of code transfer on the vehicle by relatively unskilled staff. We appreciate his suggestion of synchronous demodulation, which may be of considerable benefit, particularly since the scanning disc is now driven by an a.c. motor.

The factors which limit the range of the present scanners are the dimensions of the code-plate reflectors, their spacing and the intensity of the light source. The present maximum range is 15ft., but this could be extended if more powerful lamps were employed in conjunction with larger code strips. Lamp life is prolonged by under-running.

Mr. Bishop's analogy of lost calls with lost codes per rush hour is apt, but the number of scanners per route will most likely be based, in the ultimate scheme, not so much on the probability of missing a single bus at consecutive scanners, but rather on the probability of successive failures at any one scanner.

We are pleased that Mr. Garrard has raised the point of presenting only essential information to the operators, and agree that some form of deviation-from-normal record would be the most useful display as the system grows.

In reply to Mr. Dwivedi, we agree that additional safeguards against failure and incorrect working may be necessary. Simultaneous transmission and reception of a number of codes can be dealt with by multiplexing.

We anticipated Mr. Haigh's criticism of the motor by changing to a.c. motors early in the trials. The problem of lost or incorrect information due to double-banking and running-back is met by placing scanners at points where buses are most likely to be in forward movement near to the kerb.

621.317.79

The Institution of Electrical Engineers
Paper No. 2843 M
Mar. 1959
©

PATTERN RECOGNITION BY MEANS OF AUTOMATIC ANALOGUE APPARATUS

By W. K. TAYLOR, B.Sc., M.Sc., Ph.D., Associate Member.

(*The paper was first received* 19*th March, and in revised form* 8*th August,* 1958.)

SUMMARY

The problem of synthesizing apparatus that will automatically simulate man's ability to recognize and to learn to recognize patterns is discussed and it is concluded that analogue circuits, rather than the digital switching circuits that have been employed in the past, provide the simpler solution. A new circuit unit that possesses many of the essential functional characteristics exhibited by nerve cells in the brain is derived from earlier work on the electrical simulation of nervous-system functional activity and forms the basic element of the circuits.

The new analogue apparatus consists of a number of distinct functional circuits arranged in a definite sequence, through which signals derived from the patterns to be recognized pass simultaneously on their way to the final output terminals. Classification information may be built into the apparatus initially if it is available, but if not, it can be stored automatically in a special unit during a setting-up procedure in which samples of the pattern types that the apparatus will be required to recognize are presented, together with identification signals. Low-resolution automatic pattern-recognition apparatus is described, and examples illustrate the setting-up procedure and subsequent performance of the apparatus.

(1) INTRODUCTION

Automatic pattern-recognition problems often arise as a consequence of attempts to replace a human operator by a machine, but in the past, engineers have tended to by-pass the direct problems by introducing relatively simple manually-operated coding apparatus of their own design. Where, for example, clerks who normally received written instructions from a human source of information have been replaced by a computing machine, it has been possible to avoid the problem of designing input apparatus that will accept written instructions by introducing binary-coding apparatus that produces punched cards or tapes in response to the designer-specified manual operation of typing, performed by the human source of information in place of writing. Unfortunately, the introduction of a physically simple coding system is not always practicable, an example being the problem of letter sorting, in which it is not feasible to insist that all addresses be represented by a binary pattern of holes or black dots on the envelope. The simplicity of switching circuits[1] that are only required to recognize binary-coded patterns is due firstly to the fact that the signals derived from the patterns are nearly always above or below a chosen threshold amplitude, and this is the only 'bit' of information that the input circuit has to obtain from a single measurement. A second simplifying factor is that only $\log_2 N$ such measurements have to be made to select one of N possible binary patterns, and the points in space at which the measurements have to be made can be specified by the designer. Neither of these two features applies to handwriting, and even if black ink is used on white paper, so that the first condition is fulfilled, there are no accurately specified spatial positions for the characters as there are for the holes in punched cards or tape.

Unlike binary pattern-reading devices, the human operator is characterized by the ability to recognize an almost unlimited number of representations of the same letter, which may differ considerably in size, style or intensity and yet still contain sufficient information for the operator to make a correct selection from the alphabet. This flexibility of the human recognition system makes it necessary to consider patterns of a very general type, and it has been found desirable to define a spatial pattern as any amplitude distribution that is finite in magnitude and extent, so that it could in principle be measured by means of suitable instruments. The space occupied by a pattern may be of zero, one, two or three dimensions, but the two-dimensional case is of particular interest, since the receptive field of the human eye is effectively two dimensional. In any practical recognition system the pattern must first be detected by transducers, and it is therefore necessary to define the area or field in which patterns can be detected, and to consider only patterns that can be wholly contained within that field. Since there is also a limit to the resolving power and dynamic range of any transducer system, patterns below a certain size, or outside the amplitude range, will not be completely resolvable. With this broad definition it will be apparent that the point-by-point time average of any stationary intensity-distribution that can occur on a television screen qualifies as a two-dimensional spatial pattern. When a complete pattern is detected simultaneously by a suitable set of transducers, the outputs of the transducers constitute a secondary pattern that bears some definite relationship to the original pattern. Thus, the rods and cones of the retina and the photo-sensitive elements of a television-camera mosaic are transducers for converting light intensity into nerve impulses and electrical potentials respectively. In these examples the secondary patterns only exist as amplitudes at discrete points, whereas the original pattern may exist at all points in the field. It is thus sometimes convenient to assume that the pattern to be recognized is the distribution of voltages at a set of terminals. There is no loss of generality in making this assumption, since the voltages can represent the magnitudes of any physical variable and a continuous field can be represented as closely as desired by increasing the number of terminals corresponding to a given area of field. The simplest possible spatial pattern can be represented by a voltage at one terminal, and as the number of terminals increases so does the number of possible patterns that can occur at the set of terminals.

The number of different intensity levels that a human observer can identify after adaptation to fixed illumination is almost certainly less than ten, so that it seems reasonable to quantize the terminal voltages into levels. If the levels are L in number and exist independently at N terminals, the total number of distinguishable patterns that can occur at the terminals is L^N. In the particular cases when L is 2 or 10, the patterns of voltage may be said to be binary or decimal respectively and may be defined by writing 0 for the first level, 1 for the second level, etc., at each of the N terminals and arranging the digits in a row to form binary or decimal numbers. The numbers formed in this way can become very large for relatively simple quantized fields. If, for example, a receptive field is quantized spatially into a 9×9 matrix so that $N = 81$, and if the signal in each square of

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Taylor is in the Department of Electrical Engineering, University College, London.

the matrix is quantized into one of 10 possible levels, the total number of patterns is 10^{81}, and in theory it is possible to synthesize switching networks that will recognize any specified sets of these patterns.[1] The synthesis problem could be enormously simplified by reducing L to its smallest value, 2, but if this is done simply by letting all inputs below a clipping level be coded as 0 and all signals above this level be coded as 1, the apparatus will no longer possess the flexibility that is the important characteristic of the human recognition faculty. Switching networks that will recognize binary input patterns ($L = 2$) have been described by Uttley[2] and another author,[3] but in both cases an effective increase in the value of L would involve a considerable rise in the complexity and size of the apparatus. The analogue methods of synthesizing pattern-recognition apparatus described in the paper represent an attempt to overcome the limitations of digital methods by operating with continuous variables throughout; binary indicating signals are only generated as a final output operation.

As is the case with human recognition, the system must be supplied with classification information. This information could, in principle, by built into the apparatus by the designer if it were available in a suitable form. In practice, however, it is found convenient to arrange for the information to be stored in a special memory unit during a setting-up or 'teaching' procedure. The required classification information can then be obtained from the memory unit and used in the synthesis of a fixed-purpose machine that is similar to the learning apparatus, except for the fact that the rather complex memory unit is replaced by a simple resistor circuit.

(2) THE DEVELOPMENT OF ANALOGUE METHODS OF SYNTHESIZING AUTOMATIC PATTERN-RECOGNITION APPARATUS

The new methods of synthesizing automatic pattern-recognition apparatus avoid the difficulties associated with digital or logical methods by operating directly on the N continuous or analogue signals appearing at the output terminals of the transducers. The additional problem of deciding *a priori* which of the L^N binary input patterns of the digital system should be classified, as a particular letter of the alphabet say, is also avoided by introducing a memory device in which samples of the continuous amplitude-distribution classifications are automatically stored.

The maximum useful number of output indicating devices that can be used with the new system is 2^N, but in most applications this is reduced by discarding input signals that contain little useful information, such as signals from zones within which the amplitude is relatively constant. In addition, signal combinations representing patterns of great complexity may also be discarded, and if M output indicators remain after these operations, it is possible to classify any input distribution into one of M classes. It will be shown that all distributions placed within a given class by the apparatus are similar to those that a human operator treats as members of the same class.

The foregoing properties of apparatus synthesized by the new analogue methods are achieved by addition, subtraction, multiplication and integration operations. The new apparatus differs fundamentally from any general-purpose pattern-recognition apparatus that has been proposed in the past, since the analogue form of the signals is preserved right up to the output indicating devices.

(2.1) Basic Functional Units Employed in the New Methods of Synthesis

In all the descriptions of pattern-recognition systems known to the author, the patterns have been coded as binary signals at the input stage of the system. The basis of these earlier methods can be illustrated by considering a black and white pattern that is superimposed on squared paper. The presence of the pattern can in principle be detected, with an accuracy that is inversely proportional to the size of the squares, by means of a switching circuit supplied with on-off signals from the squares. In practice, the signals may be obtained by connecting the outputs of photo-electric cells, each arranged to collect the light from one square, to electromagnets that open or close switches when the square is more than half white. One form of switching circuit consists simply of a parallel arrangement of switches that are all open when the pattern is present but not otherwise; in other words, an 'and' gate. Unfortunately, this method is only satisfactory when the pattern intensity-distribution is well above or below the level that causes the switches to change over. In general applications the photocells give intermediate analogue-signal outputs in the range between the black and white levels, either because the white areas only partially cover some of the squares, or because the pattern itself contains intermediate intensities. If the switches change over at half the maximum white-signal level it is clear that the output of the 'and' gate will be absent if one or more signals cross the half-signal level after the gate has been constructed for an ideal black-and-white pattern. The human operator by contrast is able to recognize patterns of any intensity from threshold to a maximum, and in the case of a letter made up of dots, as in a newspaper picture, he is still able to identify the letter when a large proportion of the dots is removed at random in an almost unlimited number of combinations. The basic units employed in the new methods are designed to eliminate the high sensitivity to intensity changes that are characteristic of the switching-circuit method and to simulate more closely the human-operator characteristics. Each unit forms a signal that is proportional to the algebraic sum of a set of analogue signals derived from the patterns and is not eliminated, like the output of an 'and' gate, by insignificant pattern distortions. The new units will be called A-units; their transfer characteristics are based on the behaviour of electrical analogues of neurones in the nervous system which have been described elsewhere.[4] Each A-unit forms the algebraic sum of multiple continuously-variable input signals and delivers a proportional output signal when the sum is positive, but a zero output when the sum is negative or below a certain threshold level. The principal output of an A-unit is thus always positive or zero, but its inputs may be positive or negative. It is therefore necessary to change the sign of an output if it is required to produce a negative input to a second unit.

It will sometimes be convenient to arrange for negative instead of positive resultant signals at the input to give non-zero outputs, and for positive resultant-input signals to give zero outputs. Since the choice of sign is arbitrary, all the results derived for A-units obeying the definitions given above will also apply to A-units designed to respond to negative resultant signals if the word 'positive' is substituted for 'negative' wherever it occurs, and vice versa.

(2.2) Synthesis of A-Units

Positive and negative outputs v_{o+} and v_{o-} of an A-unit with N input terminals are defined by the expressions

$$\left.\begin{aligned} v_{o+} &= 0 \\ v_{o-} &= 0 \end{aligned}\right\} \sum_{r=1}^{N} v_r \leqslant 0$$

$$\left.\begin{aligned} v_{o+} &= k_{N+} \sum_{r=1}^{N} v_r \\ v_{o-} &= k_{N-} \sum_{r=1}^{N} v_r \end{aligned}\right\} \sum_{r=1}^{N} v_r > 0 \qquad \ldots \quad (1)$$

in which v_r is the voltage supplied to the rth input terminal. The quantities v_r are continuous variables and may be supplied by transducers that detect an amplitude distribution which has been quantized in space but not in amplitude. If an A-unit is supplied by signals from a set of transducers, its output cannot indicate that any particular set of signals has occurred. It is thus apparent that a single A-unit, like a single nerve cell in the nervous system, has little functional significance in isolation. If, however, we consider a set of A-units and compare the magnitudes of the outputs, it becomes possible under certain conditions to obtain useful information about the intensity distributions supplying the A-units by detecting the maximum output.

Two principal methods of obtaining information from the set of A-units have been employed. In the first method, the multiplying factor k_{N+} is made to depend on the number of inputs to the unit in such a way that a unit with n of its N input terminals at V volts ($n \leqslant N$) will give a smaller output than a unit with a smaller number of input terminals M and the same number n at V volts. This condition is satisfied if $k_{M+} > k_{N+}$ for $M < N$, but a second condition is necessary to ensure that the largest output is obtained from the unit having all its input terminals at V volts and also having the largest number of input terminals. These two conditions can be written

$$k_{M+} > k_{N+} \quad (M < N) \quad . \quad . \quad . \quad . \quad (2)$$

$$k_{M+}M < k_{N+}\mathrm{N} \quad (M < N) \quad . \quad . \quad . \quad . \quad (3)$$

The reliability with which the maximum output signal from a set of A-units can be detected depends on the difference between the maximum and the second largest output. A limit to the magnitude of the ratio k_{M+}/k_{N+} is set by the maximum value

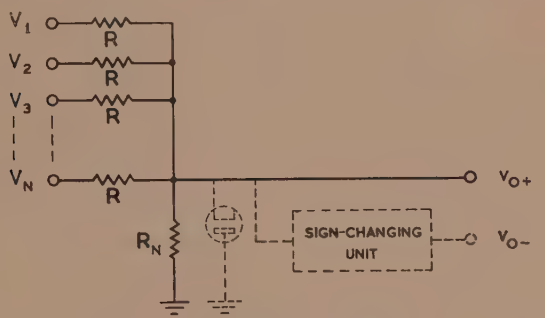


Fig. 1.—Basic circuit of A-units.

$R_N = R$ in A_1-unit.

$R_N = \dfrac{R}{N(S+1)-S}$ in A_2-unit.

of M that can occur in eqn. (3), namely $M = N - 1$. If this value is substituted in eqns. (2) and (3), the following inequality results:

$$1 < k_{(N-1)+}/k_{N+} < N/N - 1 \quad . \quad . \quad . \quad . \quad (4)$$

The multiplier k_{N+} is chosen to satisfy this inequality and, in addition, to satisfy the condition that the difference between the outputs of two A-units with N and $N - 1$ input terminals, all of which are supplied by equal non-zero signals, should be equal to the difference between the outputs of the same units when only one input of each is supplied with the same signal. This latter condition may be written as

$$k_{N+}N - k_{(N-1)+}(N-1) = k_{(N-1)+} - k_{N+} \quad . \quad (5)$$

The form of k_{N+} satisfying both conditions is

$$k_{N+} = \frac{k}{N+1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

If the constant k is unity, one possible physical realization of eqn. (6) only requires R_N in Fig. 1 to be equal to R. An A-unit with this value of R_N will be called an A_1-unit.

(2.3) Input-Output Characteristics of A_1-Units

When the value of k_{N+} given in eqn. (6) is substituted into the expression for v_{o+} [eqn. (1)] the following relationship is obtained:

$$v_{o+} = \frac{1}{N+1}\sum_{r=1}^{N} v_r \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The number of different ways of connecting the N input terminals of an A_1-unit to the output terminals of S transducers so that each transducer terminal has one or no connection is given by the binomial coefficient $\binom{S}{N}$. The number of different ways of connecting A_1-units having any number of inputs from 1 to S to the output terminals of S transducers is $(2^S - 1)$, and if the units are identified by the decimal equivalents of the binary numbers obtained by writing 1 for connection and 0 for no connection, the unit connected to all S transducers becomes number $(2^S - 1)$, while unit number one is only connected to the first transducer. If the voltage appearing at each of the S terminals is restricted for the purpose of simplifying the analysis to zero or $+V$ volts, it can be shown that the maximum output is obtained from the A_1-unit having the same identification number as the voltage pattern, the latter being defined as the binary number obtained by writing 1 for the voltage $+V$ and 0 for zero voltage at each of the S terminals. If the effects of noise are neglected, the unit giving the maximum output remains the same for all non-zero values of V and the maximum output is proportional to V. These results are derived as follows:

Let any M terminals of a set S be at $+V$ volts, while the remainder are at zero or earth potential. The output of any A_1-unit becomes

$$v_{o+} = \frac{nV}{N+1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where n is the number of inputs at a voltage $+V$ and N is the total number of input terminals. The relationships between n, N, M and S are

$$n \leqslant N, \quad n \leqslant M, \quad 1 \leqslant N \leqslant S \quad . \quad . \quad . \quad . \quad (9)$$

For each value of N the output v_{o+} is a maximum when $n = N$, and for this value of n, v_{o+} is greatest when $N = M$. The largest output, of value $MV/(M+1)$, is thus obtained from the A_1-unit with M input terminals at a voltage $+V$, and since there is only one such unit, this output must be the single maximum.

(2.3.1) Practical Example of A_1-Unit Operation.

The output voltages of a set of $(2^9 - 1)$ A_1-units, connected in all possible ways to the output terminals of 9 photomultipliers arranged in a 3×3 matrix and numbered from right to left in rows, are shown in Fig. 2(*a*) for an example in which a T pattern produces 1 volt at terminals 1, 2, 3, 5 and 8 and zero voltage at the remaining terminals. The A_1-unit giving the maximum output of 0·83 volt has five input terminals, each connected to a different one of the five S terminals delivering 1 volt. It is the unit corresponding to the binary number 111010010, which in decimal notation is 466. The identification numbers given in Fig. 2(*a*) do not correspond to the binary numbers, however, since it is more convenient to group the A_1-units according to N, the number of input terminals that they possess. The particular order shown was derived by reading the binary numbers

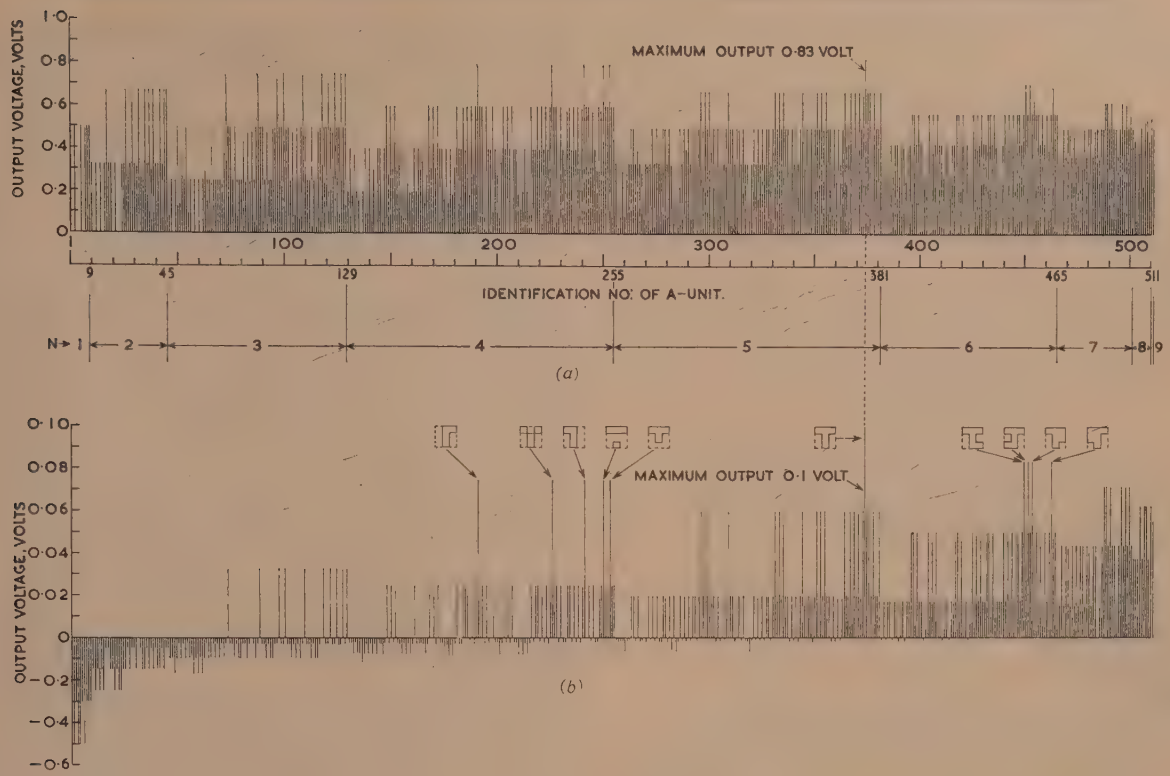


Fig. 2.—Output voltages of A-units in response to a T pattern that produces 1 volt positive signals at terminals 1, 2, 3, 5 and 8 of the 3 × 3 matrix:

1	2	3
4	5	6
7	8	9

(*a*) Output voltages of A_1-units with N input terminals in response to combinations of zero and 1 volt positive inputs. All A-units between arrowheads have N input terminals.
(*b*) Output voltages of A_2-units in response to combinations of zero, 1 volt positive and 1 volt negative inputs. Each has N input terminals supplied with zero or 1 volt positive inputs and $9 - N$ terminals supplied with zero or 1 volt negative inputs.
Ten of the 511 input patterns composed of zero or unit amplitude input voltages are shown. Arrows indicate the output voltages of the corresponding *A*-units in response to the T pattern.

from 1 to 111111111 in sequence, nine times, the first time noting all numbers with one 1, the second time all numbers with two 1's, etc., until the number with nine 1's was reached. By this means the 9 groups containing $\binom{9}{N}$ A_1-units ($N = 1, 2, \ldots 9$) were obtained in sequence and the 466th A_1-unit, giving the maximum output, occupies the 374th position as shown in Fig. 2(*a*). It will be noted that this unit also gives the maximum output for any positive value of the five equal S-terminal voltages and that the percentage excess of the maximum output $0{\cdot}83V$ above the five second-largest outputs of 0·80V remains constant at 3·7% as V is varied.

The problem of detecting the maximum will be discussed in Section 5 when it will be shown that correct indications can be obtained if the maximum A_1-unit output exceeds the second largest output by more than a minimum percentage. In general, the percentage difference decreases with n_s, the number of S terminals supplying the non-zero signals, according to the equation

$$d = 100/(n_s^2 - 1)\% \quad (n_s \geqslant 2) \quad . \quad . \quad . \quad (10)$$

The decrease in d with increasing values of n_s can be seen in Fig. 3(*a*), which shows the maximum output voltages for values of N up to 19 and for the special case of unit positive-input signals.

(2.4) Input-Output Characteristics of A_2-Units

One limitation of the A_1-unit system is the small value of d obtained when more than eight or nine of the S terminals are active. It is therefore interesting to observe that a human operator also has difficulty in distinguishing between eight or nine dots if he is not given time to count them.[5] Efforts to find a way of increasing the value of d led to the use of negative as well as positive inputs to A-units, and the way in which this can facilitate the detection of the maximum output by increasing d for a given n_s is now described.

When the N input terminals of an A-unit are connected to N of S supply terminals, there are $S - N$ supply terminals not connected to the unit. If these are connected to $S - N$ additional input terminals through sign-changing units, there will be S instead of N inputs to the unit. An A-unit employed in this manner and having the second value of R_N (Fig. 1) given by

$$R_N = \frac{R}{N(S+1) - S}$$

will be called an A_2-unit.

This expression for R_N is chosen to satisfy the condition that the output of a unit having all its N direct inputs at the same voltage and the remaining $S - N$ inputs at zero voltage shall be constant for all values of N in the permissible range $1 \leqslant N \leqslant S$. With this condition, patterns producing a large value of n_s are

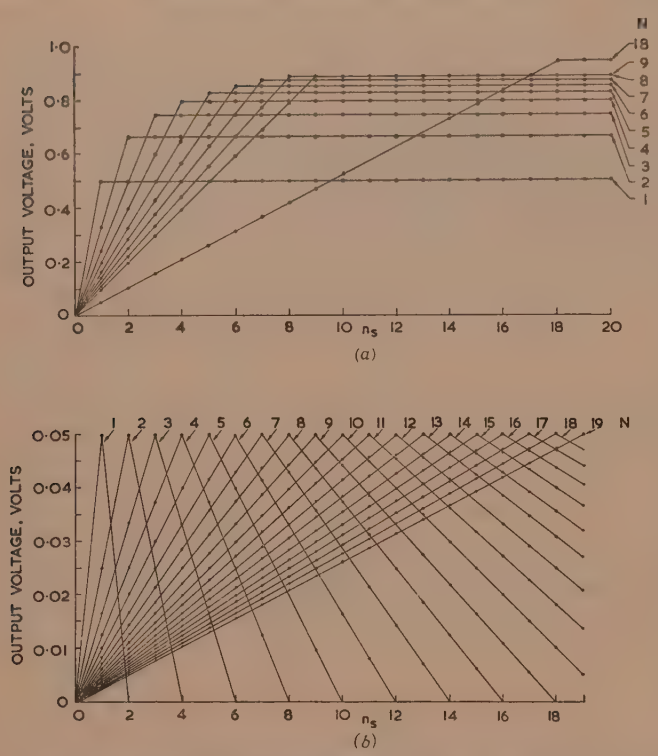


Fig. 3.—Maximum output voltages of A-units over a range of input combinations.

(*a*) Maximum output voltages obtainable from A_1-units with N input terminals connected in all possible mutually exclusive ways to S supply terminals ($S \geqslant N$) of which n_s are at a positive potential of 1 volt, the remainder being at zero or earth potential.

n = Number of input terminals supplied with 1 volt.

For maximum output of $\frac{n}{N+1}$ $\begin{cases} n = n_s \text{ if } n_s \leqslant N \\ n = N \text{ if } n_s > N \end{cases}$

(*b*) Maximum output voltages obtainable from A_2-units with N input terminals connected as in (*a*) but with S–N additional input terminals connected through sign-changing units to the remaining S–N supply terminals. Only the points correspond to output voltages in both (*a*) and (*b*).

given the same weight as those producing a small value of n_s, whereas A_1-units give greater weight to the former patterns.

(2.4.1) Practical Example of A_2-Unit Operation.

The calculated magnitudes of the outputs of a set of 511 A_2-units connected in all possible mutually exclusive ways to 9 supply terminals are shown in Fig. 2(*b*) for the particular case in which an idealized **T** pattern produces 1 volt at terminals 1, 2, 3, 5 and 8. An appreciable increase in the percentage excess of the maximum over the second largest outputs will be noticed, the numerical value of d being 24%, as compared with 3·7% when A_1-units are used. The general expression for the output voltage of an A_2-unit with N inputs directly connected and $S - N$ inputs connected through sign-changing units to S supply terminals is

$$v_{o+} = \frac{1}{(S+1)N}\left(\sum_{r=1}^{N} v_{r+} - \sum_{r=N+1}^{S} v_{r-}\right) \quad . \quad . \quad (11)$$

For the special case of unit or zero supply voltages at 9 terminals, the maximum output is 0·1 volt and is obtained from the A_2-unit with all directly connected inputs at 1 volt and all indirectly connected inputs at zero volts.

The general expression for the percentage excess of the maximum output over the second largest output when n_s of the S terminals are at equal non-zero potentials is found from eqn. (11) to be

$$d = \frac{100}{n_s}\% \quad (n_s \geqslant 1) \; . \; . \; . \; . \; . \quad (12)$$

This slower decrease of d with increasing values of n_s can be seen in Fig. 3(*b*), which shows maximum output voltages as a function of n_s for values of N up to 19. Negative output voltages, which appear at the v_{o+} terminal if the diode is not in circuit, are indicated in Fig. 2(*b*) by negative-going ordinates, but in Fig. 3(*b*) it is assumed that the diode has eliminated negative signals.

(2.5) Outputs Obtained in Response to Multi-Valued Inputs

The analysis was simplified in the foregoing Sections by restricting the supply-terminal voltages to two values, zero and 1 volt positive. All the results are clearly applicable to supply voltages of zero and $+V$ volts if inputs and outputs are multiplied by V. The response to a continuous range of input voltages at each S terminal is not obvious in general, and we therefore consider firstly the variation of any two supply voltages, v_1 and v_2, while the remainder stay at zero. When v_1 is zero and v_2 non-zero, the A_2-unit connected directly to terminal 1 gives the largest output. The reverse is true when v_2 is zero and v_1 non-zero. If v_1 and v_2 are both non-zero and equal, the maximum output appears at the A_2-unit connected directly to both terminals and will be denoted by $v_{o+}(1, 2)$. For any fixed value of v_1 there is thus a value of v_2 at which $v_{o+}(1)$ is equal to $v_{o+}(1, 2)$, the figures in brackets standing for the supply terminals to which the A_2-units giving these outputs are directly connected. This value of v_2, found by equating $v_{o+}(1)$ and $v_{o+}(1, 2)$, is equal to $3v_1$. Thus, if voltages represent light intensities, this property means that two spots of light will be recognized as two spots if the ratio of their intensities is less than three, but as one spot, the brighter of the two, if the ratio of their intensities is greater than three. It is clear that a ratio must also exist at which a human observer begins to respond to one instead of two spots as the intensity of the less intense spot is reduced towards the threshold of vision.

In general, it is theoretically possible to find the regions of an S-dimensional space within which each of the $(2^S - 1)$ A_2-units (or A_1-units) gives a maximum output, but the labour involved would be considerable for useful values of S and the results would be of little practical value. One of the important properties of the set of A_1- or A_2-units that should be emphasized, however, is that the relative magnitudes of the output voltages are constant if the relative magnitudes of the S inputs are constant. Since it is the relative magnitudes of the v_{o+} outputs that determine the final output, the apparatus gives the same indication for all patterns formed by multiplying a given intensity distribution by a positive constant at all points, assuming the extremes of overloading and poor signal/noise ratio to be avoided. This extremely useful property is not possessed by digital systems, in which intensities are amplitude-quantized into L levels, unless L is very large, which has been shown to be impracticable.

Continuous amplitude distributions that produce a well-defined single maximum at one of the $(2^S - 1)$ A_2- or A_1-units usually resemble the idealized or black and white pattern associated with the number of the unit. The infinite number of continuous amplitude distributions that can exist at S terminals are thus automatically classified by the units into $(2^S - 1)$ groups, the patterns in any selected group appearing similar to a human observer. The use of the units is thus restricted to applications in which the patterns to be recognized can be represented by two-tone drawings. This is not a serious restriction, however, since in most instances a pen and ink drawing of the light-intensity distribution that a human operator is required to recognize can also be recognized. A well-known example is the sketch of a face consisting of a few lines following the regions of steepest gradient in the original light-intensity distribution. A human observer cannot recognize small differences in the illumination of a face, but changes in the relative positions of the

regions of steepest gradient, as caused by smiling and frowning, are easily recognized. The next sections show how many members of the complete set of $(2^S - 1)$ units become redundant if the apparatus is only required to imitate this basic characteristic of human pattern-recognition.

(3) THE EXTRACTION OF RELEVANT INFORMATION FROM PATTERNS

The number of different names that we can give to patterns to indicate that we recognize them is only a very small fraction of the number of different patterns that can be resolved by eye. This is still true if only two-tone patterns are considered and the receptive field is restricted to a visual angle of 1°. Thus, the individual squares of a 5×5 mm chequerboard can just be resolved at a distance of one metre but even a well-trained human operator could not be expected to memorize different arbitrarily chosen names or numbers for the 2^{25} (or approximately 33×10^6) two-tone patterns that may be constructed by making the twenty-five 1×1 mm squares light or dark. This field contains sufficient elements for representing a large number of commonly used patterns, and, in particular, representations of the letters of the alphabet and the numerals are included in the set of 2^{25} possible two-tone patterns. It is clear that a human operator can only learn to classify a very small percentage of the total number of patterns, and this suggests that we should investigate the characteristics of the more easily recognized patterns if we are only interested in the simulation of human pattern-recognition.

(3.1) Decrease in Resolution with Increase in Pattern Size

Experiments performed by the author with the co-operation of six observers between the ages of 25 and 40 showed that the average errors in their estimation of the distance between two spots, presented at a fixed distance of 18 in, were approximately constant at 7·5% of the separation over the range $\frac{1}{8}$–$1\frac{1}{2}$ in. Thus, if the separation of two spots is r and the minimum recognizable difference in the relative positions of the spots is ds, the ratio ds/r is approximately constant. When ds is at right angles to r this leads to the reasonable result that the smallest recognizable difference in the slope of the imaginary line joining two spots is approximately independent of their separation. Further evidence in support of this is that the accuracy of estimating the position of a pointer on an unmarked instrument is not improved by increasing the size of the pointer. An automatic-recognition system employing a quantized field of equal-area squares would give constant absolute accuracy of measurement and hence improved percentage accuracy for larger distances. A more realistic field is thus one in which resolution decreases with distance from a central area of highest resolution, corresponding to the central foveal region of the human retina. For a quantized field the ratio ds/r may be replaced by dn/n, in which dn indicates the size of the quantized elements at a distance n from the centre of the field. Thus, if dn/n is 0·5, the quantized field could take the form shown in Fig. 4. To obtain the highest resolving power with this field, it is necessary to arrange for small patterns to fall on the small squares, and an automatic centring control for performing this function is described in Section 7.

In this example, the number of A_1- or A_2-units required to cover completely all possible sets of connections to the elements of the field is reduced from $(2^{1024} - 1)$ to $(2^{52} - 1)$, and providing any details can be brought within the high resolution part of the field, this enormous reduction is obtained without appreciable loss of effective resolution. The disadvantage is that only one small pattern can be inspected at any instant, so that if the field contains a number of small patterns they must be inspected in turn rather than simultaneously. This is, in fact,

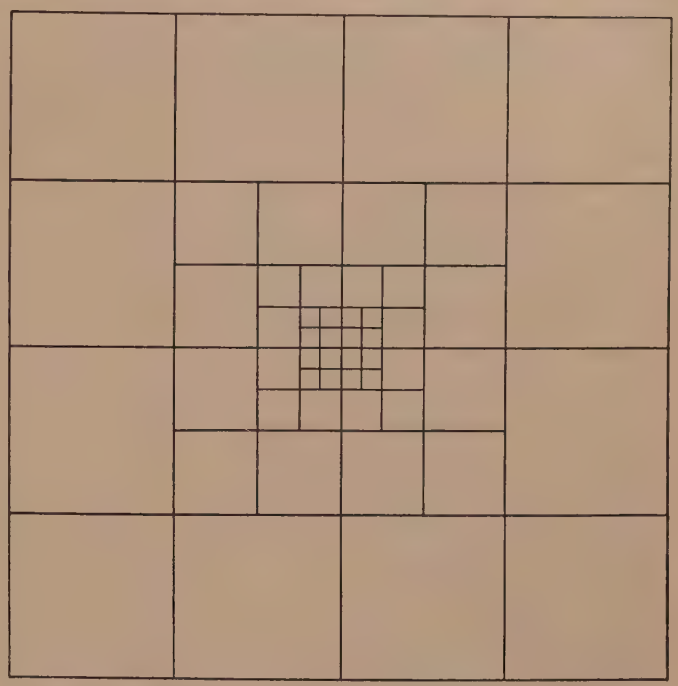


Fig. 4.—Illustration of increase in size of transducer receptive elements with distance from centre of field.

what the human operator does when confronted with a printed page, since he is unable to resolve more than five or six letters simultaneously.

In spite of the enormous reduction in the number of units obtained by grading the resolution of a specified size of field, the remaining number is still quite impracticable and the next Section shows how further significant reductions can be achieved.

(4) LIMITATION OF DETAIL

A useful measure of the amount of detail in a pattern consisting of white squares of unit side arranged on a black-background matrix measuring $n \times n$ units is the number of white-square sides that are next to black-square sides. Thus a single white square on a black background contains 4 units of detail. The pattern in which all the squares but one are white also contains 4 units of detail if the n^2 white squares round the boundary of the $n \times n$ matrix are not included. According to this definition the pattern containing most detail has alternate squares white, as in a chequerboard. The number of white square sides facing black squares is then $2n(n - 1)$. If n is even, the number of white squares in the above pattern is $n^2/2$. A pattern that contains the same number of white squares but less detail can be formed by arranging all the squares in one half of the matrix. The detail in this half black- and half white-square pattern is only the contribution of n units from the boundary between black and white, and the ratio between the detail in the chequerboard and this pattern is $2(n - 1) : 1$. The total amount of detail in the field for a specified number of white squares is thus considerably reduced if the squares are tightly packed rather than uniformly distributed, and this suggests that, if signals representing detail could be formed, they would carry less information in the former case. It is clearly necessary to take the relative positions of the parts of a pattern into account in forming such signals.

It is known[3] that a human observer is unable to discriminate between patterns consisting of more than about ten isolated dots, and if a similar limitation is placed on the automatic recognition

apparatus we are led in general to consider the outputs of A_1- or A_2-units with up to r inputs, where r is small compared with n^2. A two-tone pattern consisting of m light squares ($m > r$) on $n^2 - m$ dark squares will provide large equal inputs to all the r input terminals of $\binom{m}{r}$ units and these units will deliver equal maximal outputs. If the m squares are isolated, this set of units can be said to be overloaded with detail, but if the m squares are closely packed, say in the form of a large square, the detail will be reduced and a corresponding reduction in the ambiguity due to overloading might be expected. This potential improvement cannot be realized, however, if the relative positions of the parts of a pattern are not taken into account, since there is then no way of distinguishing between A-units receiving signals from edges and corners and A-units receiving signals from plain surfaces. To make use of the fact that the detail important for recognition is located in regions where the signal amplitude changes appreciably, it is necessary to attach greater weight to signals derived from such regions. This weighting process will be called 'detail filtering', and the detail-filter networks for performing this operation are believed to be similar to the neural networks that produce a similar effect in the nervous system. The networks were first described by the author at the Third London Symposium on Information Theory.[4] Independent work presented at the same symposium[6] described a digital computer programme for the 'edging' of binary patterns. This process resembles detail filtering but is restricted to binary variables, whereas the networks described in Reference 4 and the following Sections operate with continuous variables. The severe limitations of digital computers when applied to pattern-recognition problems are also emphasized by the fact that a computer having 65 000 bits accessible within 10 microsec took 15 min to 'edge' a single binary pattern. By contrast, the analogue detail-filter networks give a similar result in less than a microsecond and are thus a thousand million times faster, even though they can handle continuous inputs in addition to binary inputs.

(4.1) Two-Dimensional Detail Filter

The inputs to the detail filter may be supplied by the set of transducers that receives the input patterns of light intensity. If the transducers are identified by the row (m) and column (n) that they occupy in a matrix, their outputs, which are also the inputs of the detail filter, can be written $i(m, n)$. It is also convenient to arrange the input and output terminals of a detail filter in rows and columns that correspond to the transducer positions. Each filter output $o(m, n)$ is then associated with an input $i(m, n)$. A relationship between the outputs and inputs that is known from previous work[4] to give the desired characteristics is defined by the equations

$$o(m, n) = k_+ i(m, n) - k_- \sum_R o(p, q) \quad . \quad . \quad . \quad (13)$$

In the particular case for which $R = 2\frac{1}{2}$ units, the equations can be represented by the signal-flow diagram[7] of Fig. 5 in which the black nodes are signal-summation points at which the output signals are formed. At each node an input $i(m, n)$, multiplied by a gain constant k_+, is assumed to arrive from below the plane of the diagram and to add algebraically with the twenty incoming signals $o(p, q)$, multiplied by a gain constant $-k_-$, arriving in the plane of the diagram from the twenty output nodes within a radius R of $2\frac{1}{2}$ units centred on node m, n. The physical realization of the filter consists of an A-unit with appropriate resistor values at each node, twenty of the inputs being supplied by the sign-changing units of neighbouring A-units, and a single additional input by $i(m, n)$.

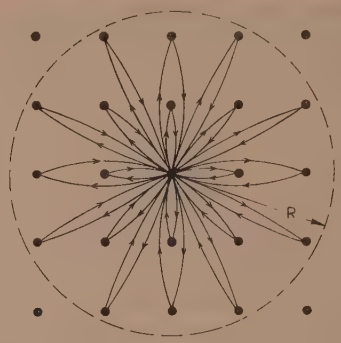


Fig. 5.—Signal flow diagram for one unit of a two-dimensional detail filter with $R = 2{\cdot}5$ units.

Eqn. (13) may be solved numerically by expanding it as the series

$$o(m, n) = k_+\{F_o[i(m, n)] - k_- F_1[i(m, n)] + k_-^2 F_2[i(m, n)] \ldots\} \quad . \quad . \quad . \quad . \quad (14)$$

in which

$$F_o[i(m, n)] = i(m, n),\ F_r[i(m, n)] = \sum_R F_{r-1}[i(p, q)]$$

If V_m is the maximum value of $i(m, n)$, it can be shown that for the given value of R,

$$o(m, n) \leqslant k_+ V_m/(1 - 20k_-) \quad (20k_- < 1)$$

The network is thus stable if $20k_-$ is less than unity. When all the inputs are equal to $+V$ the outputs are found from eqn. (13) to be

$$o(m, n) = k_+ V/(1 + 20k_-) \quad . \quad . \quad . \quad (15)$$

If, for example, the value of k_- is 1/25 and k_+ is unity, the common output voltage for equal unit inputs at all terminals is 0·55 unit. The solutions are only valid if none of the A-units have negative resultant inputs, and, assuming this condition to be fulfilled by the use of an overall background input, the output distribution in response to any two-dimensional input distribution can be calculated. Sections of the response to a square of side 21 units and unit amplitude are shown in Fig. 6. The

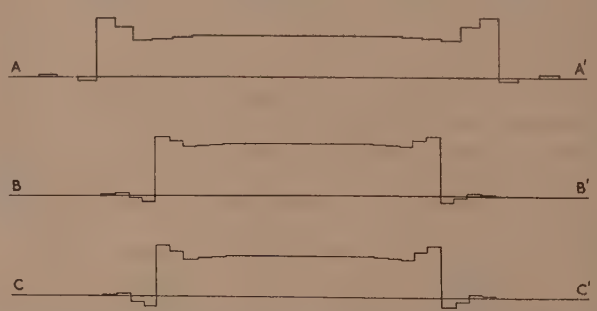


Fig. 6.—Outputs of detail filter in response to a square input pattern of unit amplitude and side 21 units.

section AA′ is taken across either diagonal of the output distribution, section B–B′ corresponds to the four edges of the square, and C–C′ to both sections through the centre of the square taken at right angles to the edges. It will be seen that the output signals are largest at the corners of the square and smallest in the central areas, but that the edges have intermediate amplitudes. The signals that contain most of the detail and shape information are thus given greater weight as required for detail filtering.

If the outputs of the set of S transducers are passed through one or more detail filters before being applied to a set of $(2^S - 1)$ A_1- or A_2-units, the outputs of the latter will be weighted according to the distribution of detail in the transducer-output distribution. For a square input pattern of constant amplitude the A_2-units with four direct inputs connected to different parts of the square will not all give equal outputs, and, if the output of the detail filter is as shown in Fig. 6, the A_2-unit connected to the four corners will give the largest output, and units connected to central areas will give small outputs. Of the units with more than four direct inputs, those connected to the corners and parts of the edges will give larger outputs than those connected to the edges or central areas. It is clear that the units giving the largest outputs are those supplied more by the corners and edges of the square than by signals from other parts of the pattern. This difference in the relative amplitudes can be used to eliminate undesirable small signals from units that are only partially activated by the input pattern. Circuits that achieve this result will be called 'maximum-amplitude filters' and are described in the next Section.

(5) MAXIMUM-AMPLITUDE FILTERS

The inputs to the maximum-amplitude filters to be described are, in general, continuous positive signals and they will be assumed to exist at N input terminals. The filters are required to operate on the signals in such a way that, if one signal i_m is larger than all the rest by more than $i_m/(1 + k_-)$, then i_m shall appear unchanged at the mth output terminal as o_m and all other output signals shall be zero. The operation of the filter may thus be represented by the equations

$$\left.\begin{aligned} o_m &= i_m \\ o_n &= 0 \qquad (n \neq m) \\ i_n &\leqslant i_m - i_m/(1 + k_-) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (16)$$

(5.1) Maximum Amplitude Filtering by Means of A-Units

The network shown in Fig. 7 will be seen to meet the requirements of a maximum-amplitude filter if reference is made to the signal-flow diagram. Inspection of this shows that, if the conditions of eqn. (16) are satisfied and all branches connected to nodes giving zero output are removed, the single non-zero output is given by

$$o_m = k_+ i_m/(1 + k_-)$$

and $o_m = i_m$ if $k_+ = 1 + k_-$ as required. The outputs of the remaining $N - 1$ units in the absence of the diodes would be

$$o_n \leqslant k_+ k_- i_m/(1 + k_-) - k_- o_m = 0$$

but since the diode in each A-unit prevents a negative output, all the o_n quantities are zero as assumed initially, and the network functions as a maximum-amplitude filter.

Signals appearing at the outputs of A_1- or A_2-units will in most cases have a single maximum, but the second largest signal i_r may be closer to the maximum than $i_m k_-/(1 + k_-)$, and if all other input signals are small enough to give zero outputs except at o_r and o_m the network equations reduce to

$$\left.\begin{aligned} o_m &= k_+ i_m - k_-(o_m + o_r) \\ o_r &= k_+ i_r - k_-(o_m + o_r) \end{aligned}\right\} \quad . \quad . \quad . \quad (17)$$

from which

$$\left.\begin{aligned} o_m &= \frac{i_m - k i_r}{1 - k^2}\left(k = \frac{k_-}{1 + k_-}\right) \\ o_r &= \frac{i_r - k i_m}{1 - k^2}(k_+ = 1 + k_-) \end{aligned}\right\} \quad . \quad . \quad . \quad (18)$$

If o_m and o_r are non-zero and

$$i_r = i_m\left(1 - \frac{1}{a}\right) < i_m\left(1 - \frac{1}{1 + k_-}\right)$$

then

$$o_r = \frac{i_m}{1 - k^2}\left[1 - \left(\frac{1}{a} + k\right)\right]$$

As an example, let $k_- = 9$, $i_r = 0{\cdot}95 i_m$ and $a = 2(1 + k_-)$. The smaller of the two non-zero outputs is $o_r = 0{\cdot}263 i_m$ and the larger output, given by

$$o_m = \frac{i_n}{1 - k^2}\left[1 - k\left(1 - \frac{1}{a}\right)\right]$$

is equal to $0{\cdot}763 i_m$. For the initial assumption of zero output at all other units to be valid, the remaining inputs must all satisfy the condition

$$i_n \leqslant \frac{k_-}{k_+}(o_m + o_r)$$

which in this example means that all other inputs must be less than $0{\cdot}923 i_m$. To take an extreme case as a second example, let all the inputs be equal. The outputs must satisfy the equation

$$o_n = k_+ i_n - k_-(N o_n)$$

from which

$$o_n = \frac{k_+ i_r}{1 + N k_-}$$

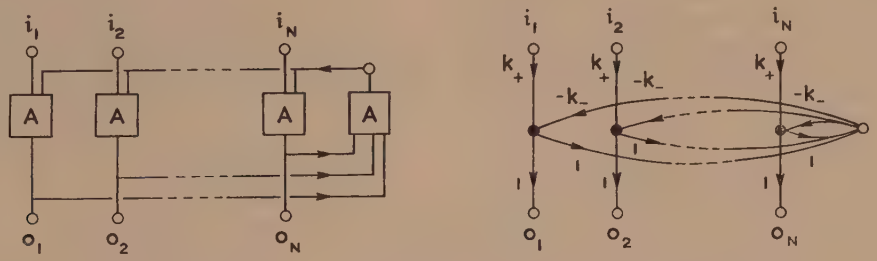


Fig. 7.—Block diagram and signal flow diagram of a maximum-amplitude filter.

For $i_n = 1$, $k_- = 9$ and $N = 10$, the value of this expression is $0{\cdot}11$ and the outputs are thus approximately one-tenth of the inputs.

Outputs that fall below a certain fraction of the inputs may be eliminated by forming a second output o_n' equal to the first output o_n minus a fraction of the input ci_n so that

$$o_n' = o_n - c i_n (o_n \geqslant c i_n) \text{ or } o_n' = 0 (o_n < c i_n)$$

Thus, if $c = \frac{1}{2}$ in the example for which $i_m = 1$ and $i_r = 0{\cdot}95$, the final outputs are $o_m' = 0{\cdot}263$ and $o_n' = 0$ for all cases when $n \neq m$. For the example in which all the inputs were equal,

the secondary outputs o_n' could be made zero by giving c the value

$$c = \frac{k_+}{1 + Nk_-}$$

It should be noted that the same method can be employed to reduce the outputs of a detail filter to zero when the inputs are all equal. In this case c must again be equal to the ratio of the output to the input. Three examples of maximum-amplitude filter operation are illustrated graphically in Fig. 8, in which the

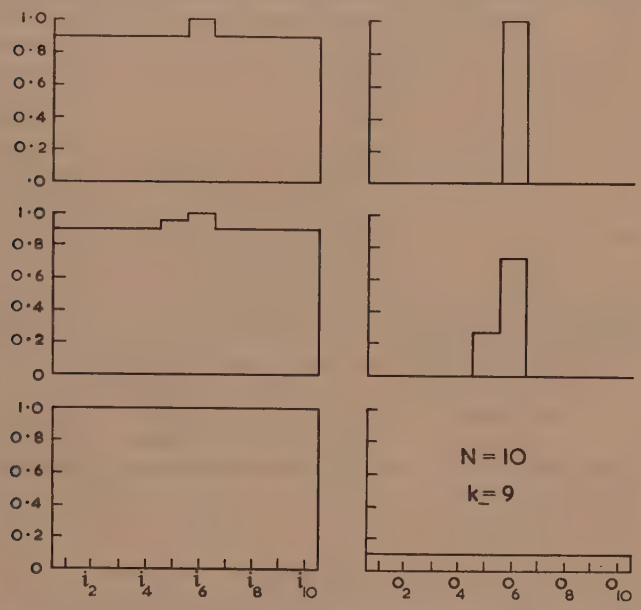


Fig. 8.—Input and output patterns illustrating the operation of a maximum-amplitude filter.

10 inputs of a filter with $k_- = 9$ are shown on the left and the corresponding outputs on the right.

It has been shown that a maximum-amplitude filter can be designed to isolate the maximum output of a set of A_1- or A_2-unit outputs, but there remains the problem of deciding how to classify each output of the maximum-amplitude filter, since in general there will be a number of outputs giving a maximum for different patterns that a human observer would classify as the same type. A solution to this problem is described in the following Sections.

may require output terminal 1 to deliver a voltage whenever a shape that would be regarded as a letter A by a human operator appears within the transducer field. An unlimited number of intensity distributions appear to satisfy this condition, however, and a detailed description of all the acceptable distributions is therefore impossible. The outputs of the set of transducers corresponding to letter A's of different shape, intensity, orientation, size and position in the transducer field are also unlimited, and when these signals have been transformed by detail filters, A_1- or A_2-units and maximum-amplitude filters, there is no obvious way of deciding how the outputs of the latter should be connected to a set of N output A-units, although the information could, in principle, be obtained by making a detailed manual analysis of the signals produced by sample patterns. The analytical difficulties can be avoided by introducing information-storage units that record the strength and duration of the maximum-amplitude filter outputs produced by sample patterns and at the same time increase the gain of the paths between the filter outputs and the inputs to the N output A-units, according to the classification signals that accompany the patterns. The sample patterns and classification signals are not necessarily supplied by a human operator. The classification signals could, for example, be generated by transducers responding to attributes of the objects that appear in the transducer field. In both cases, each classification signal is arranged to switch on a set of information-storage units connected between the maximum-amplitude filter outputs and the output A-unit that corresponds to the class of pattern presented. The arrangement will be described in more detail in Section 7.

(6.1) An Electronic Information-Storage Unit

The information-storage unit is required to multiply its present input by a quantity that increases monotonically with the time integral of past inputs, starting from the commencement of the setting-up process. By using pulse-frequency modulation followed by pulse-amplitude modulation it is possible to combine the multiplying and integrating operations in a relatively simple unit which gives an average output

$$o(t)_{av} \simeq ci(t) \int_0^t i(t)a(t)dt \quad . \quad . \quad . \quad . \quad (19)$$

The unit was first used[4] to simulate changes in synaptic efficacy in the central nervous system, and is shown in Fig. 9. The storage-capacitor section was described by Woodbury and Holdam[8] in 1949 and has since been used by Uttley[2] to store the conditional probabilities of binary patterns. The arrange-

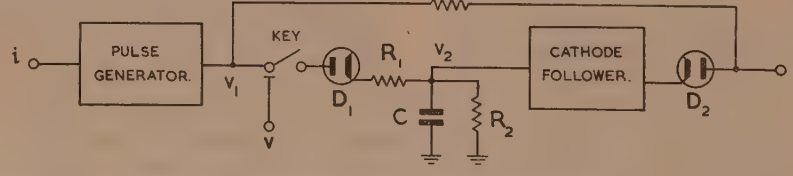


Fig. 9.—Electronic information-storage unit.

(6) AUTOMATIC INFORMATION STORAGE

For practical reasons the number of transducers must be limited, but there is no limit to the number of different patterns of intensity that may occupy the field covered by the transducers. The number of output terminals must also be limited, and if there are N of these, each input pattern may either produce an output at one terminal or no output. The designer must first decide on N; he could then attempt to define the patterns that should produce an output at each terminal. If, for example, the patterns to be classified are all possible letter A's, the designer

ments described in these two papers do not give the desired function, however, since they do not include the pulse-frequency-modulation and pulse-amplitude-modulation sections of the new unit.

In the circuit in Fig. 9, the voltage waveform at v_1 consists of a train of pulses of width t_p, height V and a 'frequency' $f(t)$, defined as the reciprocal of the interval between adjacent pulses occurring near time t. If this interval is denoted by $T(t)$ we may write

$$f(t) = 1/T(t) \simeq k_f i(t) \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Each pulse increases the charge on C through R_1 and if leakage is neglected, the voltage across C after n pulses is

$$v_2 \simeq \frac{Vt_p}{CR_1}n \quad (nt_p \ll CR_1) \quad . \quad . \quad . \quad . \quad (21)$$

The number of pulses charging C in a time interval t, during which the key may be open or closed according to whether the value of $a(t)$ is 0 or 1, is given by

$$n \simeq k_f \int_0^t i(t)a(t)dt$$

and since the output-pulse height is limited to v_2 by the diode, the output waveform is

$$o(t) \simeq v_1(t)\frac{t_p}{CR_1}k_f \int_0^t i(t)a(t)dt$$

The average amplitude of the pulse train waveform $v_1(t)$, taken over an interval in which the pulse frequency is practically constant, is $k_f i(t)Vt_p$ and on making this substitution we obtain

$$o(t)_{av} \simeq \frac{k_f^2 t_p^2 V}{CR_1} i(t) \int_0^t i(t)a(t)dt \quad . \quad . \quad . \quad . \quad (22)$$

which is the required form. The averaging operation corresponds to demodulation of the frequency- and amplitude-modulated pulse train $o(t)$ by means of a low-pass r.c. filter.

(7) EXPERIMENTAL AUTOMATIC PATTERN-RECOGNITION APPARATUS

The signal-flow graph of a complete but relatively simple experimental pattern-recognition system of low resolution is shown in Fig. 10. The inputs to the nine detail-filter terminals

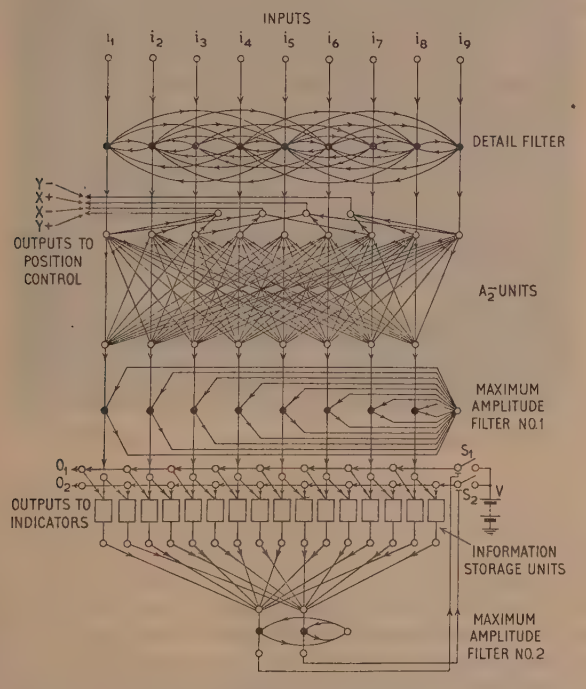


Fig. 10.—Signal flow diagram of experimental automatic pattern-recognition apparatus.

are supplied by nine photo-electric multiplier (p.e.m.) units arranged in the form of a 3 × 3 matrix to collect light reflected from patterns. The value of R for the detail filter is 1·5, which means that the signal-flow graph for each A-unit is identical with that shown in Fig. 5 when all except the nine central nodes and their connections are omitted. Values of the constants k_+ and k_- defined in Section 4 are 19 and 0·4 respectively.

The number of A_2-units in the experimental apparatus is restricted to eight by applying the principles discussed in Section 4. The eight units are those satisfying the conditions that N should be three and that none of the inputs should be adjacent. If this latter condition is applied to other values of N we find that nine units are required for $N = 1$, sixteen for $N = 2$, one for $N = 4$ but none for higher values of N. If patterns are centred automatically by combining outputs of the detail filter to form position-control signals X_+, X_-, Y_+ and Y_-, it is found that sixteen of the A_2-units with $N = 1$ and $N = 2$ correspond to unstable positions and can therefore be discarded.

There are two maximum-amplitude filters in the apparatus, the values of k_+ and k_- being respectively 2·3 and 3·2 for the first filter and 20 and 5 in the second filter. None of these values is critical; the values of k_+ were chosen to give suitable signal-amplitude ranges and the values of k_- were the smallest that prevented more than one output indication occurring at any instant of time. The second filter ensures that only the largest sum of eight information-storage-unit outputs reaches the final output stage. The outputs of the first maximum-amplitude filter supply eight pulse generators and these in turn supply sixteen capacitor storage units. In general there would be $R \times T$ storage units for R A_2-units and $T(\leqslant R)$ output terminals, and in the present case $R = 8$ and $T = 2$. The values of the components shown in Fig. 9 are $R_1 = 100$ kilohms, $C = 12\,\mu$F, and $R_2 = 100$ megohms.

During the setting-up or teaching process the switches S_1 and S_2 in Fig. 10 are closed either manually or by means of electromagnets driven by externally generated classification signals. When the apparatus becomes automatic the switches are closed by electromagnets supplied by the outputs of maximum-amplitude filter No. 2 as indicated in Fig. 10. In both cases the closed switch supplies a final output voltage V to the corresponding output terminals O_1 or O_2. The same voltage V is also connected to an input of the information-storage units (Fig. 10) that corresponds to the terminal marked V in Fig. 9. This voltage closes the key and allows pulses generated by the input i from maximum-amplitude filter No. 1 to charge the storage capacitor C.

When the p.e.m. matrix is required to centre on and track moving patterns, the signals X_+ and X_-, obtained by adding outputs 3, 6, 9 and outputs 1, 4, 7 of the detail filter, control the current flowing in the split field of a d.c. motor so that the resultant flux is proportional to $(X_+) - (X_-)$. These two sets of three signals will be seen to correspond respectively to the right and left halves of the 3 × 3 p.e.m. matrix when it is remembered that the p.e.m.'s are numbered in rows from left to right. If X_+ is greater than X_- the horizontal motor drives the matrix in a clockwise direction, and vice versa. Similar conditions apply to the control of the vertical motor by the signals Y_+ and Y_-, with the result that objects are fixated and tracked automatically. Unfortunately it was not possible to use this facility in the following experiments, since portions of the asymmetrical patterns employed fell outside the small receptive field in the stable position. This would not occur with a larger matrix.

(7.1) Results Obtained with the Experimental Apparatus

It should again be emphasized that the experimental apparatus has low resolution for reasons of economy, and that it will only recognize patterns falling within a restricted size range. Furthermore, there are only two outputs, which means that only two classes of shapes can be recognized. The response to a selection

of **T** shapes will be considered first. An enlarged negative of a typewritten T was used as a pattern in the first half of the experiment. Initially the storage capacitors were all discharged, and on holding the T in the centre of the receptive field there was no output from the second maximum-amplitude filter and the switches S_1 and S_2 remained open. This condition is illustrated at position A_1 in Fig. 11. In each position the upper number

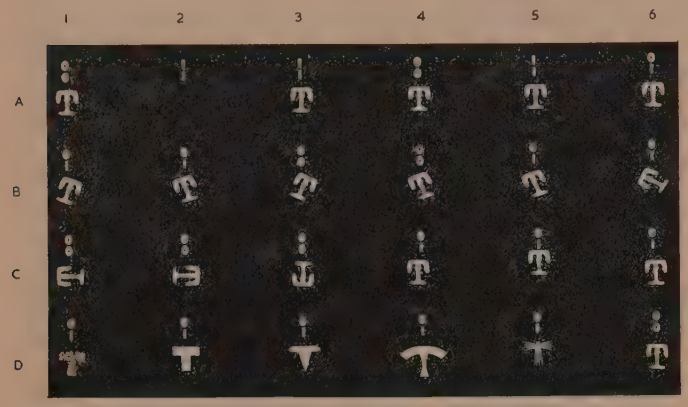


Fig. 11.—Response to **T** shapes before, during and after setting-up procedure.

indicates that the switch with the same number is closed manually, and the lower number indicates automatic closing of the corresponding switch and hence the automatic output of the apparatus. The lower number is unconditionally the same as the upper number if the latter is non-zero as shown at A2. When the upper number is zero, i.e. when there is no classification input, a non-zero lower number indicates automatic closing of the corresponding switch and hence the automatic output of the system. The patterns were photographed at the time of presentation and are shown below the numbers, the centre of the receptive field being indicated by a small black or white dot.

In setting up the apparatus, S_1 was closed manually for a few seconds and then released (A3, Fig. 11). This was repeated until the switch remained closed when the manual pressure was released (A6), an electromagnet supplied by the first output of the second maximum-amplitude filter providing the closing force. The **T** pattern could then be said to be recognized automatically as a member of class 1 pattern types and the voltage V appearing at terminal O_1 could be used to control any desired form of indicator or to actuate a typewriter key. The tests from B1 to D5 were designed to determine how much change in position and shape could be tolerated before the patterns fell out of the **T** class. The typed T was recognized when tilted through an angle of approximately 20°, as shown at B1 and B2, but if the angle was increased slightly beyond this point there was no response (B3 and B4). The angle of tilt for recognition could, however, be increased by slowly rotating the figure through intermediate angles until at B6 an angle of 60° was tolerated. This effect is made possible by the overlapping of the p.e.m. unit fields and the consequent transfer of light to different sets of units as the pattern moves. If the movement is sufficiently slow, a new set of storage capacitors is charged before the original pathways cease to carry signals, with the result that the switch is held closed. A pattern classified in one position may thus give the same output in other positions and at different magnifications if a sufficient number of intermediate samples is presented. At C1 the **T** pattern was presented horizontally and gave no response, but it should be noted that horizontal **T**'s could be classified as the same pattern type as vertical **T**'s by closing S_1 manually during this presentation. This would be useful, for example, if we were interested in automatically sorting **T** shapes from a random distribution of shapes with all possible orientations. Positions C4, 5 and 6 show that small displacements of the **T** from the centre of the field are tolerated by the apparatus. Five patterns that were found to give the same response as the original typed T are shown at D1–D5. All the foregoing presentations were made over a period of 45 minutes, the memory being reinforced by each successful recognition. The last presentation, D6, was made half an hour after D5, and by this time the storage capacitors had discharged below the levels required to produce closure of the switch. If, however, a permanent type of information-storage unit had been employed, the final **T** would have been recognized after any lapse of time.

Permanent recognition apparatus for fixed applications may, of course, be constructed by measuring the gains of the information-storage units after setting-up is completed, replacing each unit by a resistor attenuator giving an attenuation proportional to the gain. By this means the memory set up in general-purpose apparatus as the result of 'instruction' may be transferred to simpler fixed-purpose apparatus. The simplification may be considerable in apparatus containing a large number of storage units, since any paths of zero gain that remain after the setting-up procedure may be omitted from the fixed-purpose apparatus.

The simultaneous storage of pattern information from a number of different shapes leads to independent classification when the shapes are sufficiently unlike one another, as are the five letters shown in Fig. 12. These letters may be classified in

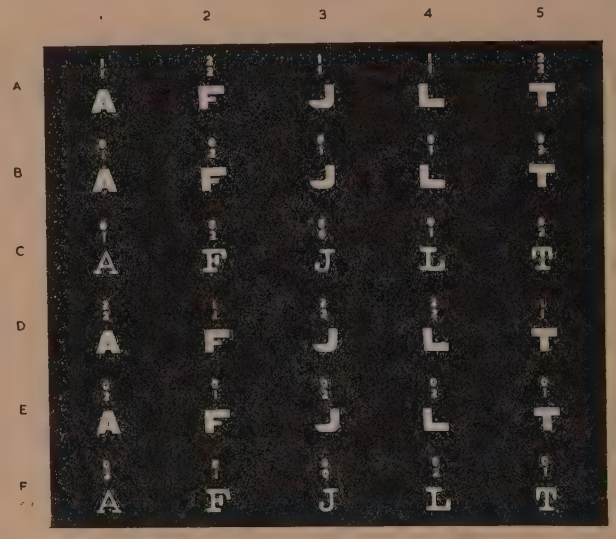


Fig. 12.—Response to a set of random classifications.

any one of $3^5 = 243$ different ways, since each may be of class 0, class 1 or class 2. Initially all patterns are in class zero and they may or may not be moved to a non-zero class. If zero classification is ruled out, the letters may be classified in any one of $2^5 = 32$ ways and one of these, chosen at random by spinning a coin, is shown by the top row of classification numbers (1, 2, 1, 1, 2) at A in Fig. 12, which indicate the number of the S-switch to be closed when the letter below the number is presented during the setting-up procedure. Row B indicates the automatic-recognition outputs obtained after the operations of row A. Without any additional classification inputs, four of the different style letters of row C were recognized correctly, but the **J** was too far removed in style from the first **J** and did not give an indication. As a further test of generality the storage capacitors

were first discharged and the setting-up procedure was repeated with the 1's and 2's of the first experiment interchanged, as in row D. The results, apart from the number changes, were the same as before and are shown in rows E and F.

The foregoing examples illustrate the simplicity of the setting-up procedure and it will be noted that this remains the same no matter how large the apparatus. A sufficiently large machine could presumably be set up to recognize human faces by closing a different key for each face presented but the same key for samples of a given face in different aspects and at various distances. It is obvious that an analytical approach to this problem would be quite impracticable and that the automatic pattern-recognition apparatus provides a simple solution. Unfortunately, the complexity of the analysis problem in the case of elaborate patterns precludes the synthesis of apparatus to give a specified accuracy and the only alternative appears to be an experimental approach in which the size is increased until the required performance is attained.

(8) CONCLUSIONS

Automatic pattern-recognition apparatus synthesized according to the methods described will simulate many characteristics of man's faculty for learning to recognize visual patterns. By operating directly with continuous analogue variables the apparatus avoids many of the coding problems associated with digital switching systems. Although it is not possible to predict the storage capacity that would be required to match the human faculty, the results suggest that apparatus of practicable size would recognize the alphabet and numerals in a variety of styles and sizes and whether typed or handwritten. Should this prove to be the case, it would add further support to the theory[4] that networks similar to those described in the paper exist in the visual system of the brain and that A-units and information-storage units are respectively functional analogues of the neurones and synapses of the brain.

(9) ACKNOWLEDGMENTS

The author wishes to thank Professor H. M. Barlow for his interest and helpful advice during the research described in the paper.

He is also indebted to Professor J. Z. Young for many stimulating discussions on functional analogues between biological and electrical systems.

The financial support of the Nuffield Foundation and the assistance of the College technical staff are also gratefully acknowledged.

(10) REFERENCES

(1) 'Synthesis of Electronic Computing and Control Circuits' (Harvard University Press, 1951).

(2) UTTLEY, A. M.: 'The Conditional Probability of Signals in the Nervous System', Radar Research Establishment Memorandum No. 1109, 1955.

(3) 'Electronic Reading Automation', *Engineer*, 1957, **203**, p. 414.

(4) TAYLOR, W. K.: 'Electrical Simulation of Some Nervous System Functional Activities' ('Information Theory', Edited by Colin Cherry, Butterworths Scientific Publications, 1956), p. 314.

(5) HUNTER, W. S., and SIGLER, M.: 'The Span of Visual Discrimination as a Function of Time and Intensity of Stimulation', *Journal of Experimental Psychology*, 1940, **26**, p. 160.

(6) SELFRIDGE, O. G.: 'Pattern Recognition and Learning' ('Information Theory', Edited by Colin Cherry, Butterworths Scientific Publications, 1956), p. 345.

(7) MASON, S. J.: 'Feedback Theory—Some Properties of Signal Flow Graphs', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 1144.

(8) WOODBURY, R. B., and HOLDAM, J. V.: 'Counting' ('Waveforms', Edited by Chance, Hughes, MacNicol, Sayre and Williams, McGraw-Hill, 1949), p. 615.

681.142 : 518.5

The Institution of Electrical Engineers
Paper No. 2792 M
Dec. 1958
©

A SYSTEM FOR THE AUTOMATIC RECOGNITION OF PATTERNS

By R. L. GRIMSDALE, M.Sc., Ph.D., Graduate, F. H. SUMNER, B.Sc., Ph.D., C. J. TUNIS, B.Eng., M.Sc., Ph.D., and T. KILBURN, M.A., Ph.D., D.Sc., Member.

(*The paper was first received* 12*th August, and in revised form* 4*th October*, 1958. *It was published separately in December*, 1958.)

SUMMARY

The paper describes a new method for the automatic recognition of patterns. The method may be applied to any form of spatial pattern, but in the present instance, patterns consisting of line figures are considered. The pattern is presented to a flying-spot scanner connected to a digital computer. The shape of the pattern is analysed and a statement is prepared describing the basic features of the pattern. The pattern is then recognized by comparing this statement with a number of others already stored in the computer which relate to named patterns. Patterns are recognized independently of the angle at which they are presented to the scanner, and may be of any size provided that limits imposed by the resolution of the scanner are not exceeded. The average time to recognize a character is 60 seconds with the system programmed on a medium-speed computer. Special-purpose equipment built to perform certain of the stages of the process, together with the use of higher-speed computers now envisaged, will reduce this time by at least a thousandfold.

If a new pattern is presented to the machine it will indicate its inability to recognize the pattern, but by giving the machine the name of the pattern, it may become one of the standard patterns which it can subsequently recognize. All the patterns recognized by the machine are hand-drawn and consist of such symbols as the capital letters of the alphabet and numerals, although the system is in no way limited to any special set of characters. Using exactly the same method but with an increase in the degree of complexity, it will be possible for machines to read handwriting.

Special allowances are made for imperfections in the patterns, including breaks and general ill-definition. Where there is some confusion and an unknown pattern resembles two or more of the standard patterns, the relative degrees of similarity of the unknown to each of these standard patterns is printed out by the machine.

(1) INTRODUCTION

Automatic recognition of characters—printed or hand-written—is becoming increasingly important. The paper describes a system for automatic pattern recognition. The distinction between a 'character' and a 'pattern' is important because both a Gothic and an Arabic A may be considered to be the same character but are entirely different patterns. A less extreme example of this is the symbols 4 and 4.

The system at present is restricted to line figures, although it could be adapted to solid ones. The pattern is fed into the computer using a flying-spot scanner and is represented within the machine as a two-dimensional array of points in the store (Fig. 1). The pattern is examined by a programmed scan which starts as a television-type scan until part of the pattern is found. The scan then changes to follow the particular part of the figure, the scan progressing essentially in a vertical direction. This mechanism produces a description of the part of the figure, giving details of the length, slope and curvature. This part is then removed and descriptions are obtained of the remaining parts. A further process assembles the parts of the figure as scanned into more basic sections which are to form the 'statement'. The statement is then compared with the statements of the patterns 'known' to the machine. A considerable degree of flexibility is included in this recognition section, and the statement referring to the unknown pattern need not be identical with a particular statement of a known pattern, but need only be nearly identical.



Fig. 1.—Computer console display of stored pattern.

The system can thus produce a statement for an unknown pattern which subsequently becomes one of the known statements. In this way the machine can 'learn' new patterns.

The present system differs from the majority of methods[1–9] which have been devised or proposed for character recognition, because it obtains a description of the shape of the figure. The majority of the other methods involve a geometrical comparison of the unknown figure with a set of standard figures. The simplest one[3] uses a rotating disc and an optical system with a photocell to measure the amount of agreement. It is not always realized that with such systems the amount of disagreement between the unknown and the standard figures is just as important as the amount of agreement. This is particularly evident in the comparison between the symbols C and O.

Other methods,[5–8] including the present one, involve some sort of scanning, either by a vertical bank of photocells moving

Dr. Grimsdale, Dr. Sumner and Dr. Kilburn are in the Electrical Engineering Department, University of Manchester.
Dr. Tunis was in the Electrical Engineering Department, University of Manchester, and is now with I.B.M. Endicott Laboratories, U.S.A.

horizontally, or by a flying-spot scanner and a single photocell. Such scanning systems are employed for economic reasons, although a two-dimensional array of photo-sensors, as in the human eye, would give higher speeds.

The use of figures printed with magnetic ink[9] is being extensively developed in America. The figures are magnetized and passed under a single magnetic head. The number of symbols which can be distinguished by this method is severely limited, to about 12. The figures must be of a special form and differ somewhat from normal typescripts, particularly if a small degree of rotation of the figure is to be allowed.

One system[7] uses a flying-spot scanner with vertical scanning lines. As the scanning point passes over the figure, the position, number and length of the intersections of the figure by the scan are recorded. The figure is recognized by details of these intersections.

Another system involves the use of a flying-spot scanner to divide the pattern area into a matrix of 10 by 10 points.[6] The information (black or white points) contained in this matrix is used to activate a set of logical circuits. These test both for the concurrence of observed black areas with those of a standard symbol, and for the concurrence of white areas. Displacement of the character is allowed for by the use of a 'pre-scan', which positions the scan relative to the character. The number of symbols handled is limited, and rotation of the characters is not tolerated, but the method is very fast.

One of the difficulties inherent in the process of scanning symbols printed on paper is the poor quality encountered in



Fig. 2.—Enlarged typewritten character.

practice. Fig. 2 shows an enlarged symbol printed by a normal typewriter. When the blurring and mutilation which the symbol may suffer in its subsequent history are superimposed, it becomes apparent that any processing of the scanner output must take account of these imperfections. This problem is considered in the paper; it has also been considered in a different way at M.I.T.[10] Another approach to the problem of pattern recognition is one in which a conditional probability machine is used.[11]

(2) GENERAL DESCRIPTION OF THE SYSTEM

The operation is performed in a number of stages, each of which will be considered in detail. The pattern is first transferred from paper to the store of the computer. The pattern is then examined by a programmed scan. The result of the scan is to produce descriptions of segments of the figure, i.e. divisions which are conveniently produced by the scanning process, and are not necessarily true divisions. The vertical and horizontal lines comprising a letter **L** might be regarded as true divisions. The scanning process also includes measures to allow for figure imperfections, dirt on the paper, and other forms of 'noise'.

In the 'assembly' part of the programme which follows, the segments of the figure obtained in the scan are analysed and connected, wherever appropriate, into true figure parts; a description is given of the length and slope of straight-line parts and the length and curvature of curved portions. The scan and assembly sections of the programme together produce the 'statement', which gives a complete description of the figure. This description is independent of the orientation and size of the figure, the lengths of the various parts being given relative to one another. It is, in effect, a coded representation of the pattern. It may be regarded as a one-dimensional pattern which consists of symbols chosen from a restricted range. For example, there is one symbol to represent a straight-line portion and another to represent a curved portion. In the 'recognition' section which follows, a comparison is made between the one-dimensional pattern or statement describing the unknown pattern and the statements already stored within the computer, together with the names of the patterns they describe. A considerable time would be wasted if every stored statement had to be examined; to prevent this, an automatic classification system is provided; this examines the stored statements and arranges them into classes according to common features. As an example, all statements describing figures with, say, four ends would be grouped together; then, at an early stage in the recognition process, a statement referring to an unknown pattern which had four ends would be compared with the statements which are members of the 'four ends' class.

The criterion for a successful comparison is that there must be a certain measure of agreement between the 'unknown' statement and the statement referring to the pattern with which it is to be identified. This must be substantially greater than the agreement between the unknown statement and any of the other stored statements. The relative importance of the features embodied in the statements, used in assessing the measure of the agreement, may be stated by the programme designer or may be acquired by the system as the result of 'experience'. In this case it will be necessary to measure the time to search for the required stored statement, using various classification systems, and to give a high score to those which yield a minimum. On subsequent searches, classification systems with a high score will be used first.

(2.1) The Scan

The object of this part of the scheme is to produce a form of geometrical description of the figure; such a description will not be ideal, but will provide sufficient and convenient information for the assembly programme to process into the statement. One of the difficulties with the scanning of an actual pattern is that

the quality is often very poor and may undergo further deterioration in the transfer to the computer using the flying-spot scanner. The other difficulty is that, to achieve greatest speed, it is best to programme the computer to perform what is basically a horizontal scan. Vertical or nearly vertical lines may then be readily examined, but horizontal lines present more difficulty. It would be desirable to provide special facilities in the computer or in some special-purpose equipment for easier vertical scanning. The whole of the pattern is examined, although this amounts to examining the edges after testing that the particular horizontal portion of the pattern has an adequate amount of 'black' or figure points between the edges.

The Manchester Mark I Computer has been used for this work and so it has been convenient to choose a raster size of 40×64. The picture area is then represented in the machine by 64 forty-digit lines. Points on the pattern appear as points on these lines; this is illustrated in Fig. 1. Scanning is conveniently done from right to left along the horizontal lines of this stored pattern area, and from bottom to top. The scanning programme operates by analysing the pattern into 'groups'. Broadly, a group is defined as a two-dimensional collection of pattern points, the horizontal extent of which displays no sudden changes on successive lines. Examples of groups are shown in Fig. 3. In (*a*), group 1 consists of the lower vertical

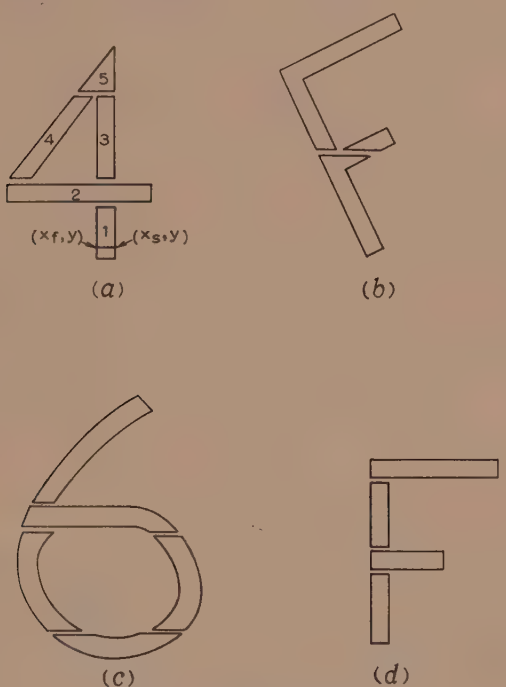


Fig. 3.—Examples of groups.

portion. Group 2 is taken as a new group because the horizontal length of its sets of points is much greater than the corresponding length for group 1. Similarly, groups 3 and 4 are separately identified. The junction of these two forms a new and distinct group, No. 5. Further examples of group formations are shown in (*b*) and (*c*). It is clear that the groups into which a pattern is analysed by the scan will depend on the orientation of the pattern with respect to the scanning lines. Fig. 3(*d*) shows the group formation of the pattern at (*b*) for a different orientation. However, the assembly programme which follows will yield identical statements for (*b*) and (*d*).

For purposes of description it is convenient to refer to a particular portion of a pattern on one scanning line as a 'section'. In Fig. 3(*a*), the section is completely described by the co-ordinates of its extremities, (x_s, y), (x_f, y).* There may, of course, be more than one section per scanning line, but each will belong to a different group.

The scanning programme has a number of significant features:

(*a*) *Noise Figure.*—This allows for the absence of points from any particular section, due to irregularities or imperfections of the pattern. It also prevents the inclusion into a group of spurious points due to dust or dirt on the pattern or to flying-spot scanner noise.

(*b*) *Interval and Gap Counts.*—When a particular section has been found, it is tentatively assumed that there will be similar sections on lines above. However, owing to imperfections there may be a complete, or almost complete, absence of points on one or more of the lines; such gaps must be tolerable, and for this reason the gap count is provided. The interval count is similar to the gap count in that it allows for the presence of an excessive number of points on one or more of the lines. The interval and gap counts are related to the noise figure, and this is set automatically by the programme to one of three values according to whether the pattern is considered to be small, medium or large. As a future development, the noise figure will be adjusted automatically if the imperfections of the figure result in its being divided into too many groups.

(*c*) *Quick Look.*—This is a section of the scanning programme which simulates a form of curve following. By the use of existing machine logical functions, it enables a simple group to be quickly followed; it is based on the assumption that the position of a section on the next line to be considered will be predictable from information of an existing group. When complications appear, such as the intersection of two groups, this section of the programme will be abandoned for a more elaborate examination in the section of the programme known as 'detailed scan'.

(*d*) *Detailed Scan.*—The features of this part of the programme are illustrated below.

(2.1.1) A Sample Scan.

The action of the scanning programme and the formation of groups is illustrated here with reference to Fig. 4. The noise figure, the allowable gap count and the allowable interval count are all assumed to have the value 2. Scanning begins on the bottom line ($y = 0$) under the control of the 'master' section of the scanning programme and proceeds from right to left.

Lines on which the number of points is less than the noise figure are ignored, and when the first line to be considered is found, lines below this one will not be referred to again. The first line to be examined here is $y = 11$, and it is examined point by point for pattern points. This results in a description of the section as $x_s = 24$, $x_f = 20$ and $y = 11$, and these values are recorded.

The section is presumed to be a member of a possible group, and if further sections of this group are to be found, they will lie on the line above in roughly the same range. Accordingly, the points $x_s + 5$, $x_s + 6$ and $x_f - 5$, $x_f - 6$ on the next line are first examined. If there are no pattern points at these positions, the quick-look section of the programme is entered; otherwise the slower detailed-scan section is used.

The quick-look routine examines only the points between $x_s + 5$ and $x_f - 5$ on the line above, and the beginning and end of the section on line 12 are rapidly found.

A section is defined as a collection of pattern points on a line which is continuous, i.e. which has no gap greater than the

* It is implied that the space between x_s and x_f is filled with pattern points. However, in view of the provision of the noise figure (q.v.), it is not essential that every digit position between x_s and x should contain a point.

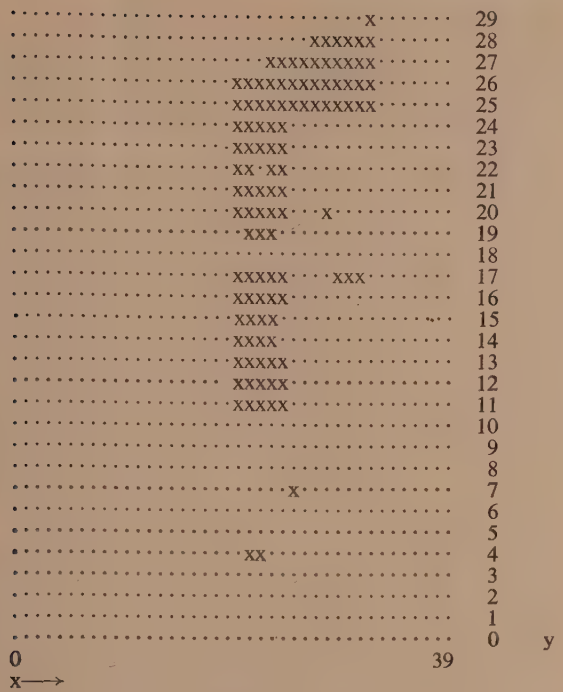


Fig. 4

noise figure; thus the section on line 12 is defined by $x_s = 24$ and $x_f = 20$. This new section is now compared with the previous section of line 11, to establish whether both are members of the same group. The tests used are given below.

(*a*) *Overlap Test.*—Two sections on consecutive lines are members of the same group only if the range (x_s to x_f) of one includes the range of the other.

xxxxxxxx xxxxxx xxxxxx xxxxxx xxxxx xxxxx
(i) (ii) (iii)

Thus the sections at (i) and (ii) are accepted as members of the same group while those at (iii) are not.

(*b*) *Length Test.*—The length ($x_s - x_f$) of a proposed section is compared with that of the last member section of the group, and the section is accepted if its length is within a given range of the member section.

(*c*) *Edge Continuity Test.*—If two sections are to be members of the same group, at least one of the edges must be continuous.

xxxxxxxxxxx xxx xxxxxxxxxxx xxxxx
(iv) (v)

Thus sections at (iv) are not accepted as members while those at (v) are. The situation at (v) arises for straight lines making a small angle with the direction of the scanning lines.

The noise figure in use is taken into account in all the tests, and all tests must be successful for a new section to be accepted as a member of the current group.

The sections on lines 11 and 12 (Fig. 4) are clearly members of the same group, as are the sections on lines 13, 14, 15 and 16, and this is established by the quick-look routine. However, the presence of the points $x = 29$–31 on line 17 is noted when the points $x_s + 5$, $x_s + 6$ are examined as a preliminary to the quick-look routine; as a result, the detailed-scan routine is entered instead of the quick-look routine.

Detailed scan considers the points at $x = 29$, 30, and 31 as a complete section since the nearest neighbour on the line ($x = 24$) is more than two points (= noise figure) distant. This section fails to qualify as a member of the current group, being rejected by the overlap test. The remainder of the line is then examined for further sections. That between $x = 20$ and $x = 24$ is found and accepted as a group member.

As each member section is accepted, its co-ordinates (x_s, x_f) are entered in a list. In addition, the difference between these co-ordinates and those of the previous section are calculated and retained until a new section is found. Thus the (x_s, x_f) for the last two sections are always available.

No points are found on line 18 and a gap count of 1 is noted. The values of x_s, x_f and the co-ordinates for line 18 are interpolated, using the current x_s, x_f values and the co-ordinates of the last section member of the group. Thus, the gap is temporarily filled in, and the quick-look routine can examine the correct range on line 19 for further sections. That on line 19 is accepted and the gap count is reset to zero. If, however, no sections belonging to the group are found on lines 19 and 20, the gap count will be increased to 3, and since this is larger than the allowable value, the group is deemed to be terminated.

The presence of the point at $x = 28$ on line 20, being more than 2 points (= noise figure) from its nearest neighbour, results in the quick look being abandoned, and a detailed scan is performed on that line. The point ($x = 28$) is rejected as a possible group member, and the remaining points are examined and accepted as a section belonging to the current group, as are the sections on lines 21–24.

The section on line 25 is rejected by the length test as a member of the current group. An interval count of 1 is recorded. The interval count operates in a similar manner to the gap count and is successively increased to 2 and 3 on lines 26 and 27. The value of 3 exceeds the allowable interval count, and the group is terminated. The master section of the scanning programme is then re-entered.

At this time the group is described by a list of x_s, x_f values preserved in the high-speed store of the computer, and by the lines on which it begins and ends. A test for triviality is first applied, to determine whether the group is worth 'remembering' as a component of the pattern. The criterion for rejection is the size of the group, and the test is included to eliminate stray groups not belonging to the pattern, such as ink spots which are too large to be ignored by the application of the noise figure. The current group is accepted, and its points are removed from the pattern to prevent reconsideration.

The group data must now be averaged to remove small irregularities in the edges of the group. The smoothing thus produced is necessary for a later stage in the information analysis, when the values of x_s, x_f are effectively differentiated to find the slopes and curvatures of the group edges. The averaging process consists in replacing the individual values of x_s by average values of x_s taken four at a time. This is, of course, also done for the x_f values. The averaged and unaveraged edges of a group are shown in Fig. 5. Groups with less than four lines cannot be treated in this way and are labelled to receive special attention; groups with less than six lines are also labelled because they too require special treatment in later stages.

The group considered in this sample scan is, however, averaged since it is 14 lines high. The total data pertaining to it are now stored away and the co-ordinates of its beginning and end sections (x_s, x_f, y_b), (x_s, x_f, y_e), together with the position in which it is stored, are noted in a group directory.

The scanning process recommences on line 11 and the section $x_5 = 29$, $x_s = 31$ is found on line 17. This is tentatively assumed to be a new group. Failure to find further section members on lines 18–20 results in this group being deemed trivial.

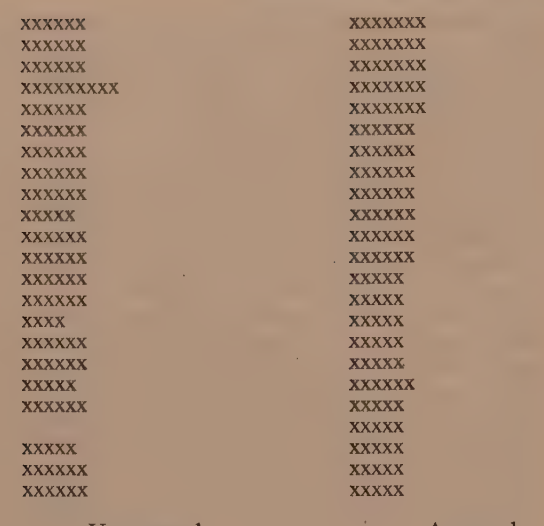


Fig. 5

Scanning begins again on line 17 and the section on line 25 is found and assumed to be the beginning of another group. The process is continued until all sections have been removed from the pattern area.

(2.2) Assembly

It is the purpose of the assembly programme to process the information concerning the individual groups obtained from the scan programme and to produce a simple coded representation—the statement—describing the pattern. This is done in two parts; in the analysis section a description is obtained of the lengths, slopes and curvatures of the various groups comprising the pattern; the synthesis section, which follows, determines the relationship between the individual groups and prepares a table of joins.

(2.2.1) Analysis.

The particular nature of the scan programme produces two types of groups which it is convenient to identify separately. These are horizontal and non-horizontal groups. A horizontal group is one which has its length directed along the scanning lines. The sections at the ends of these groups (i.e. at the top and the bottom of the groups) may be rejected as members of the group, by virtue of the length test; a special examination is therefore performed, and those adjacent horizontal groups which overlap sufficiently are combined and redescribed as a single horizontal group.

The remainder of the description of the analysis programme applies essentially to non-horizontal groups, but, of course, where appropriate, a similar treatment is given to horizontal groups.

The scan programme prepares a table of values of x_s and x_f, the starting and finishing abscissae for each section of the groups. The x_s and x_f values are treated separately at first, and since the processes are identical, only the treatment for x_s values is given. The column of values in the x_s column is differenced to obtain the Δx_s values and these are subsequently differenced to obtain $\Delta\Delta x_s$. The values of $\Delta\Delta x_s$ are now smoothed; this proves necessary because, for patterns with irregular edges, the action of effectively differentiating twice accentuates these irregularities, and the general tendencies become masked.

A further column is produced giving values of the cumulative second differences, $\Sigma\Delta\Delta x_s$. This new column, at any line, represents the total change of slope of the group edge from the first line of the group to the present line. For a straight-edged group, the entries in this column will, ideally, all be zero; for a curve, the successive entries will display either a gradual increase or decrease.

When the nature of the trend of the cumulative-sum column changes, there will have been a change in the nature of the group edge, say from curve to straight line. In order to detect these changes, two further columns are formed. These are headed Last Increase and Last Decrease and represent respectively the number of lines since the last increase in the cumulative-sum column and the number since the last decrease. When the nature of the group edge changes, a zero entry will be made in one of these columns, in which the entries have previously been increasing steadily. The change is considered significant if the value preceding the zero is greater than a value b, where b is some fraction of the total length of the group. Such changes serve to divide the group edge into 'edge-sub-elements', which are separate straight-line or curved portions of the edge.

The next stage is one of correlating the edge-sub-elements obtained for the two edges of a group, to form its 'elements'. This is done in the following manner.

Using their end co-ordinates, the length and overall slope of each sub-element may be calculated. The lengths of the first left-hand edge and right-hand edge sub-elements are compared and their difference in length (ΔL) is noted. If the right-hand edge-sub-element is longer, its length is compared with the combined lengths of the first two left-hand edge-sub-elements. If the difference in length is now larger than ΔL, the first edge-sub-element on the right is compared in other respects (slope and curvature) with the first edge-sub-element on the left. If the difference in length is smaller than ΔL, however, a tentative combination of the first three left-hand edge-sub-elements is made, and the process of length comparison is repeated. Thus the edge-sub-elements of either edge are matched with respect to length before a more detailed comparison is made. Further, the matching is such as to attempt to combine the sub-elements of one edge, if the other edge has a simple description.

The slope and curvature of the matched sub-elements of the two edges are now compared. If these are in general agreement, an element of the group is formed. The average length, slope and curvature of the two edge descriptions, as given by the sub-elements, is recorded, together with the end co-ordinates (x_s, x_f, y_b), (x_s, x_f, y_e) of the element.

If the slopes of the matched sub-elements do not agree, any combined sub-element is broken down into its components and these are individually tested for agreement with the slope and curvature of the sub-element of the opposite edge. If agreement is obtained, group elements are formed; if not, 'abnormal elements' are produced. Examples of abnormal elements are shown in Fig. 6; these are elements (*a*) 2, (*b*) 2 and (*c*) 1. Fig. 6(*b*) illustrates a case where the two sub-elements of the right-hand edge are tentatively combined in order to match the length of the single sub-element of the left-hand side. But the overall slopes disagree and the right-hand side is broken down into its components, sub-elements 1 and 2.

If the matching sub-elements agree with respect to overall slope, but not with respect to curvature, no attempt is made to break down the combined sub-elements (if indeed, there are any) and an abnormal element is formed. Fig. 7 illustrates a group with edges of differing curvature. This case will be detected by the slopes comparison and thus described as two elements, one of which, 2, is abnormal.

Abnormal elements due to different edge slopes are described by the average length of the two edges and by both edge slopes.

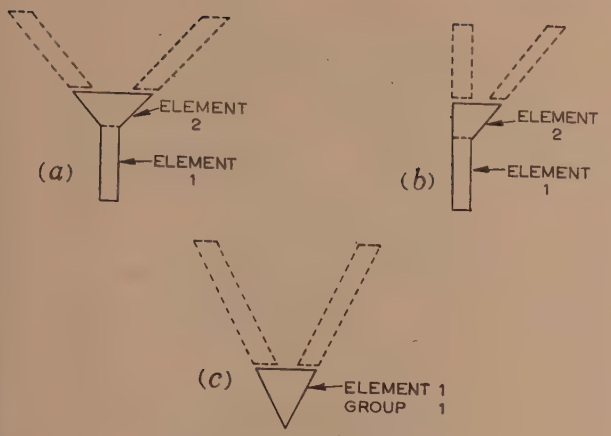


Fig. 6.—Abnormal elements.

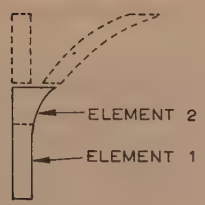


Fig. 7

These are used at a later stage in joining groups to one another, when slopes must be matched. Abnormal elements due to different curvatures are described by the average length, slope and curvature of the two edges involved. All abnormal elements are specially labelled for future identification.

2.2.2) Synthesis.

The groups of the pattern are now described in terms of their elements. The present section determines the relationship between individual groups, and a table of their joins to one another is prepared. Some of the various joins that can occur are shown in Fig. 8 and a sample table of joins for the *R*

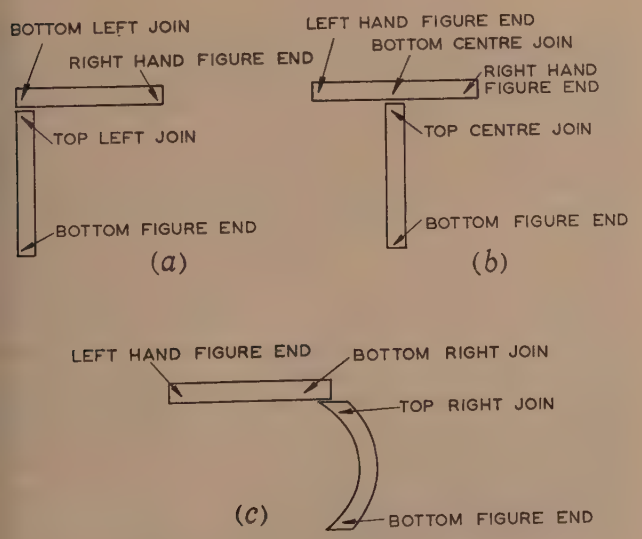


Fig. 8.—Types of joins and ends.

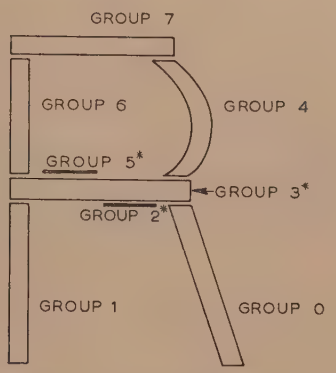


Fig. 9.—Groups of an R pattern.
* Groups 2, 3 and 5 combined and designated as 'Group 2'.

	BT	TB	TT	BB
	R L C	R L C	R L C	R L C
Group 0 ..	.. – – –	2 – –	– – –	– – –
Group 1 ..	.. – – –	– 2 –	– – –	– – –
Group 2* ..	.. 0 1 –	4 6 –	– – –	– – –
Group 3* ..	.. – – –	– – –	– – –	– – –
Group 4 ..	.. 2 – –	7 – –	– – –	– – –
Group 5* ..	.. – – –	– – –	– – –	– – –
Group 6 ..	.. – 2 –	– 7 –	– – –	– – –
Group 7 ..	.. 4 6 –	– – –	– – –	– – –

* Groups 2, 3, 5 are combined and designated 'Group 2'.

BT bottom-top join.
TB top-bottom join.
TT top-top join.
BB bottom-bottom join.

R right join.
L left join.
C centre join.

The numbers in the columns refer to the adjoining groups.

Fig. 10.—Table of joins.

pattern of Fig. 9 is given in Fig. 10. The four columns of the table of joins correspond to:

Joins from the bottom of the current group to the top of another (BT joins).
Joins from the top of the current group to the bottom of another (TB joins).
Joins from top to top (TT joins).
Joins from bottom to bottom (BB joins).

There is a row for each group and the numbers of the groups to which it joins are in appropriate columns. These signify whether the join concerned is to the right, left or centre, as illustrated in Fig. 8.

It may be seen that provision is made to record only a limited number of joins. This restricts the current application to relatively simple patterns, but it is not a fundamental limitation of the method used.

When there is no adjoining group at one end of a given group, this group is said to have a 'figure end'. The nature of these ends is determined, and the types of ends possessed by a group are denoted by a label in the group directory. Types of ends are shown in Fig. 8.

With the group connections established, the group elements may be examined in detail to eliminate those which have been introduced by the interaction of two groups or the presence of figure ends. Types of these elements are illustrated in Fig. 11. Groups which consist entirely of abnormal elements, such as that shown in Fig. 12, are also examined further. These will carry labels signifying bottom or top intersections as a result of the previous classifying process. If these groups do, in fact, join to two other groups (appropriately), they themselves are effectively removed from the directory, and the two adjacent

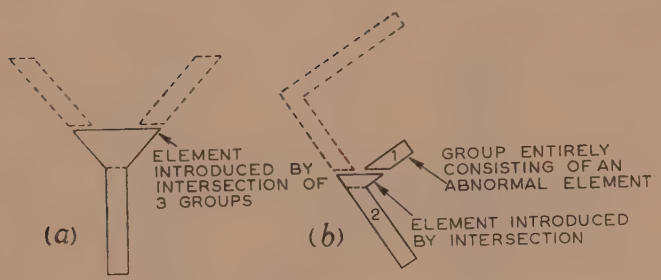


Fig. 11

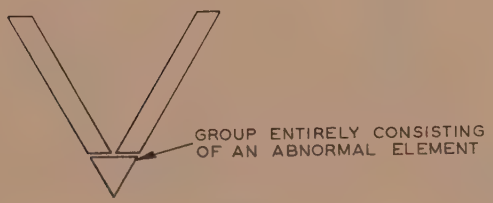


Fig. 12

groups are denoted as joining each other in the table. Group 1 in Fig. 11(*b*) will also consist entirely of an abnormal element. It is not a true intersection, however, since it joins only one other group of the pattern. The final element of group 2 is an abnormal one and thus both its edge slopes are preserved. This applies equally to group 1. With the information concerning the nature of the join of group 1 to group 2, appropriate slopes may be compared to verify a proper match. If this is successful, group 2 is redescribed as a straight-line group having a figure end and a slope equal to the average matched slopes.

The information concerning abnormal elements is not required after the group connections have been made and groups obscured by figure ends have been reconstructed. Accordingly, abnormal elements are removed from the group descriptions, their lengths being added to those of adjacent elements, as part of a general condensation process which is performed on the group descriptions. Elements with similar slopes and curvatures are combined and any remaining trivial elements are removed.

The various groups comprising the pattern have been analysed and described and their interrelations noted. However, these groups are a consequence of the internal scanning process of the machine and do not necessarily represent the simplest description of the pattern. For example, the pattern of Fig. 13(*a*) is broken up into three groups with group 1 having a top centre join to group 2 and group 3 having a bottom centre join to group 2. A simpler description of this pattern is obtained if groups 1 and 3 are combined to form a new group which crosses group 2.

The object of the main part of the synthesis section of the programme is to form the 'segments' of the pattern, segments being defined as the components of the simplest description of the pattern. Since groups are not always identical with pattern segments, some combinations of the groups may be required. Further examples where this is necessary are given in Figs. 13(*b*)–(*e*).

Broadly speaking, the synthesis takes place in two phases. The first is synthesizing groups which have been separated owing to intervening or adjoining groups. The second is synthesizing groups which are members of curved segments and which have been separated owing to the nature of the scanning process. Examples of the first type are shown in Figs. 13(*a*), (*b*), (*c*) and (*e*), and of the second in Fig. 13(*d*).

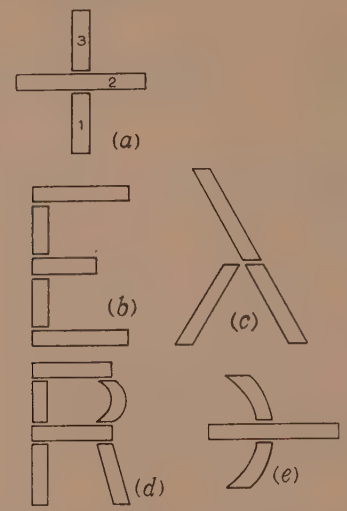


Fig. 13.—Examples of division into groups.

The next process to be carried out is the compression and coding of the information concerning the segments of the pattern and their interrelationship, to form the statement.

The slopes ascribed to curved segments are determined in a convenient manner; these remain unaltered, as the direction they define is unique with respect to the remainder of the pattern. This is not the case for 'horizontal' groups, as is shown in Figs. 14(*a*) and (*b*). The line at (*a*) will have a slope of 170°,

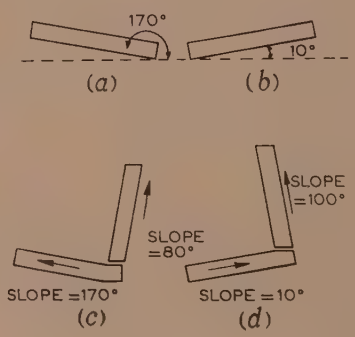


Fig. 14.—Definition of slope.

while that at (*b*) will have a slope of 10°, yet they may be the same line rotated through 20°. The result will be that, if the segment is part of a larger pattern [Figs. 14(*c*) and (*d*)], its relative slope with respect to other segments will be radically changed by rotating the pattern slightly. This is not the case, however, if the slope of the segment is determined by the direction of travel from its end—if it possesses one. It is thus convenient at this stage to redefine the slopes with respect to the end wherever possible; this serves an additional purpose, as will be seen from consideration of Figs. 15(*a*), (*b*) and (*c*). For if each of the lines of the patterns shown has a slope as determined from their ends, the type of joins they possess, i.e. whether left or right, top or bottom, need not be specified.

This cannot, of course, apply to straight-line segments having no figure ends or two figure ends. With the latter, allowance is made in the recognition process to tolerate 180° differences in the slopes of these particular segments.

The slopes of all segments, which have hitherto been in degrees to permit detailed comparisons, are adjusted to have one out of 32 possible values. The total length of the pattern

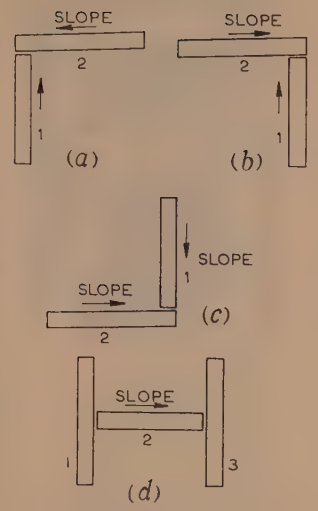


Fig. 15.—Definition of slope using figure ends.

s calculated, and the length of each segment is normalized again o have one of 32 values. The entire data concerning the pattern egments are now coded into 40 binary digits. This is the repre- entation used at present, and it will suffice only for relatively imple patterns such as the English alphabet capitals and umerals. A small increase in the coded data will allow the nique description of more complex patterns. The 40-digit oded representation gives the length, slope and curvature of the egment, together with details of the ends and joins to other egments.

The set of 40-digit representations forms the statement for the attern.

(3) RECOGNITION

The main object of the present work is the recognition of atterns, but it is anticipated that the method will be applied to haracter identification; the distinction between pattern recog- ition and character identification has already been mentioned n the Introduction—it is that several patterns may be used to epresent the same character. The present Section is primarily oncerned with pattern recognition.

The machine has available a list of standard pattern statements hich is extended as the machine learns new patterns. The roblem involved in pattern recognition consists essentially of electing one of the standard pattern statements which agrees ufficiently well with the statement relating to the unknown attern. Two systems have been devised for performing this peration, the second being an improved and more versatile ethod; both have been used and they are now described, with ome suggestions for further systems.

(3.1) Recognition—System 1

The first stage is to determine the number of segments in the nknown pattern statement. A reference segment is now chosen ut of those forming the unknown pattern statement. This egment should preferably be one which does not involve any mbiguity in the slope. Thus it should be a straight-line segment hich has one end as a figure end; failing this, a curved segment chosen; if neither of these is available, a straight-line segment ith two or no figure ends is selected. In this latter case a label attached to the reference segment to indicate the possibility f a 180° ambiguity in slope. The reference segment is the one which the slopes of all the other segments of the unknown pattern are referred so that the pattern may be described independently of its particular orientation.

A corresponding segment must now be found in the standard pattern statement selected from the stored list. To do this a number of tests are applied; these tests, which are applied to the reference segment and each of the standard pattern segments, allow

(*a*) No difference in the number of figure ends or joins.
(*b*) A difference of up to one quadrant in the specification of a curved segment.*
(*c*) No difference in the number of elements.

After these tests have been satisfied, a further test is applied which allows a difference in length between the two segments proportional to the length of one of the segments involved. If no suitable segment is found in the standard pattern, another standard pattern is chosen. When the segment corresponding to the reference segment is found, the difference in slope ΔS_p between the two segments is determined. An attempt is then made to match each segment in the unknown pattern with each segment of the standard pattern. To do this, the above tests are applied, together with a test to compare the slopes of the corresponding segments in which due allowance is taken of the ΔS_p value. If these conditions are fulfilled, segments in the standard pattern statement are rearranged and renumbered to match the unknown pattern statement. When this is done, the joins of the segments are compared and these must agree. That is, if group 2 of the unknown has a side join to group 1, the same must apply to groups 2 and 1 of the standard pattern. If the joins agree, the unknown pattern has been identified and the 'name' of the successful standard pattern is printed out. If the joins do not agree, a new standard pattern is considered, and if the pattern cannot be identified with any of the standard patterns the machine will indicate failure and await further instructions. If required, the unknown pattern may become a new standard pattern.

(3.2) Recognition—System 2

Recognition system 1 is limited in scope and requires complete agreement at certain stages. System 2 has a greater degree of flexibility and uses an improved directory system to give rapid selection of the standard pattern statements.

After the statement for the unknown pattern has been found, as at the end of Section 2.2.2, a 40 digit 'key word' for the pattern is formed. The key word is made up of 8 blocks of 5 digits giving in turn the numbers of segments, straight-line segments, one-segment curves, two-segment curves, 3 or more segment curves, ends, side joins and crosses.

In the first stage of the recognition process of system 2, the key word for the unknown pattern is formed and compared with all the key words for the standard patterns, and the degree of agreement with each is noted.

The statement whose key word gave the maximum score is now compared in detail with the statement for the unknown pattern. The unknown pattern statement is first modified, if necessary, so that it has the same number of segments as the standard pattern. All such modifications are noted, as they affect the final score. If these modifications are so severe that an excessive adverse score is incurred, the particular standard pattern is rejected and another chosen.

All standard patterns which have obtained more than a certain score are now compared, one by one, in greater detail with the unknown pattern. One segment of the unknown pattern is chosen; this should be preferably the longest straight-line segment having one figure end, or, failing this, the longest curved line. This segment is compared with the first segment

* Curved segments are specified in terms of the number of quadrants they occupy.

of the standard pattern. A score is obtained which is based on the agreement of number of ends, side joins, crosses and length. If the score is sufficiently high, the difference in slope is determined and recorded. The remaining segments of the unknown and standard patterns are compared in all possible combinations using the difference in slope already found. A total score of agreement is accumulated. This is now repeated taking the same original segment of the unknown pattern with the second segment of the standard pattern. Again the difference in slope is calculated, and the total scores for all the other segments taken with each other in the various possible combinations are determined.

The whole procedure is repeated with all the patterns whose key words showed reasonably high agreement with the unknown pattern key word.

If one of the standard patterns shows a sufficiently large amount of agreement with the unknown, and this agreement is much greater than the agreement of the unknown with any other standard pattern, the name of the best standard pattern is printed out; otherwise the machine indicates that there is some confusion between two or more patterns. As with system 1, there is the facility that if the pattern is not recognized it can be made into a new standard pattern.

(3.3) Future Developments

It is possible to envisage a number of interesting systems which could be included in future recognition schemes. One of these may be used in the recognition process when applied to character identification. When a large number of standard pattern descriptions are stored within the machine, time may be wasted in choosing suitable standard patterns to be compared with the unknown. Accordingly a classification system more detailed than that described in Section 3.2 would be desirable.

It is intended that such a classification system will develop after a large number of standard patterns have been recorded in the store. The significant features are then determined. These are the features which are special to some but not all patterns; for example, classes might be formed of patterns with 4 ends or patterns with 3-segment curves. The unknown patterns are examined to find which significant features they possess; the appropriate class is then selected, and a more detailed examination is performed to differentiate between the members of the group.

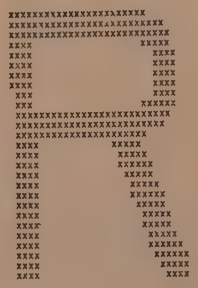
A further development to facilitate both the classification process and to improve the speed of the system is to employ some special-purpose equipment to perform the scanning of the pattern on the paper. The scanner should be capable of performing certain special operations like curve-following under direct control of the computer. One of the parts of the programme which occupies a large amount of time is the scanning procedure which determines the slope and curvature of the parts of the figure. The computer at present examines the pattern in rather too great detail, and it would be preferable if the scanner could indicate if there were more than a certain number of picture points in a certain region. In this way the scanner could indicate general trends.

(4) CONCLUSIONS

The system described has been tested and has operated successfully on a large range of patterns. The computer is first supplied with a library of standard patterns consisting of all the capital letters of the alphabet and the numerals. These standard patterns are drawn in white on a black background, a typical example being shown in Fig. 16. Each standard pattern is presented to the scanner in the vertical position, and the



(*a*)



(*b*)

Fig. 16.—Pattern reproduction.
(*a*) Original character as presented to the scanner.
(*b*) Digital representation of character.

computer obtains a statement which describes the essential features of the pattern. The computer is given the name of each standard pattern as it is presented, and it stores the statement for each pattern together with its associated name in its library. Owing to the limited resolution of the flying-spot scanner, the patterns which the internal scanning system of the machine operates with are not so well defined as in Fig. 16(*a*) but appear as in Fig. 16(*b*). It is patterns of this form which the machine has to process in the recognition system.*

Having been given a library of standard patterns, the machine is capable of recognizing all of them when subsequently presented, independently of the orientation. Shifts in the vertical or horizontal directions, provided the pattern remains in the field of the scanner, also present no difficulty. The computer prints the name of the pattern, the difference between the angle at which this pattern is presented and the angle at which the standard pattern was presented. When the presented pattern resembles two or more standard patterns, the names of these are printed and with each there is a number to indicate the degree of agreement or certainty.

Using the same library of standard patterns, the machine is capable of recognizing a range of other patterns. A group of these is shown in Fig. 17. They have the same general form as the standard patterns, but not their precise shape. Fig. 17(*e*) illustrates a case where the machine indicates uncertainty. The pattern could either be the letter T, with the join not central, or a poorly drawn letter L. When this pattern is presented, the machine prints the following:

```
L  320   30
T  140   45
```

The figures following the letters give, in the first column, the angle of the pattern from the vertical measured in degrees in

* These figures have been prepared by the computer using a teleprinter; they are copies of the two-dimensional array which appears in the store at the beginning of the process.

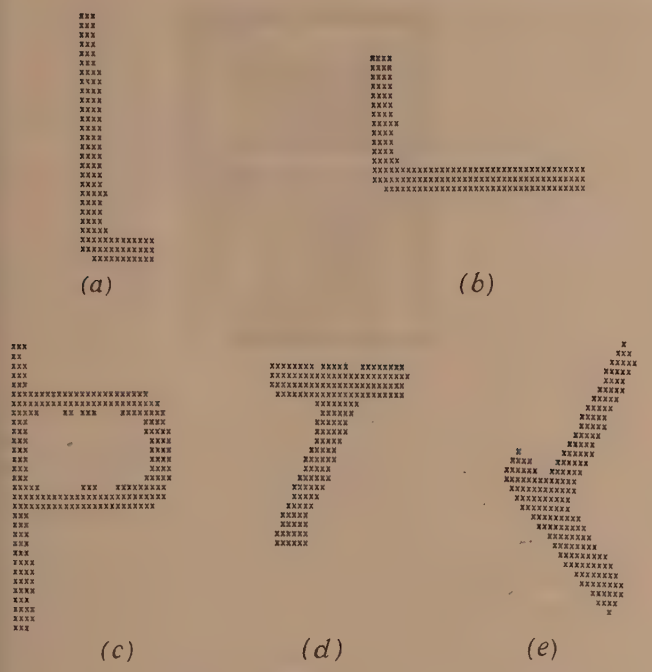


Fig. 17.—Deformed characters which can be recognized.

he anti-clockwise direction, and in the second column, the measure of certainty of the particular answer. If great significance is placed on ends in the scoring system used by the recognition part of the system, as occurred in the present case, the **T** is the more probable pattern. If, however, greater importance had been attached to the angles of rotation of the patterns, the **L** answer would have been the more probable.

The system can also deal with patterns of various thicknesses (Fig. 18). The thin-line patterns are simple to analyse and recognize, but thick-line patterns present difficulties to recognition machines in general. The present system successfully recognizes thick figures, provided they can be adequately resolved by the scanner. It is a feature of the internal scanning mechanism that the whole of the figure is examined. Whereas information is only recorded about the edges of the figure, all the space between the edges is scanned, and it is implicit in the information given about the edges of the figure that the figure is solid. Thus when the **F** pattern shown in Fig. 18 is scanned, it is broken into segments in the same way as for a thin-line figure. The system

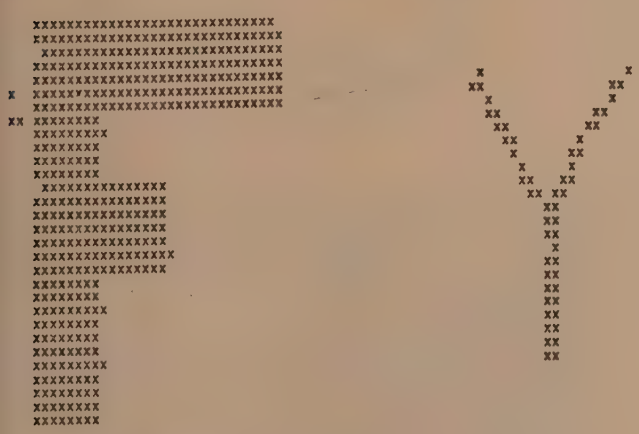


Fig. 18.—Examples of thick and thin characters.

as at present arranged fails to obtain correct statements for a few curved figures at certain angles, because there are insufficient scanning lines to determine the curvature, as illustrated in Fig. 19. Nevertheless, when an incorrect description has been obtained, the recognition programme still succeeds in giving a name to the pattern in certain cases. This difficulty can be overcome either by increasing the resolution of the scanner or by a process of first reducing the thick figure to a thin one. The effect of such a reduction is shown in Fig. 20.

The patterns need not be perfectly regular, and various imperfections such as breaks, ragged edges and dirt on the paper are tolerated. The bad patterns in Fig. 21 are all recognized, again using the library of standard patterns for comparison. There are two reasons for this recognition: first, there is an averaging process and mechanism for ignoring breaks in the pattern, controlled by the noise figure (Section 2.1); secondly, the information is reduced to statements describing lines and curves. There is thus a preliminary recognition procedure in the scanning process, in that straight and curved lines are recognized from ill-defined patterns.

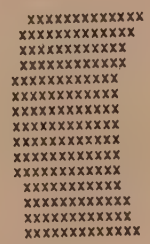


Fig. 19

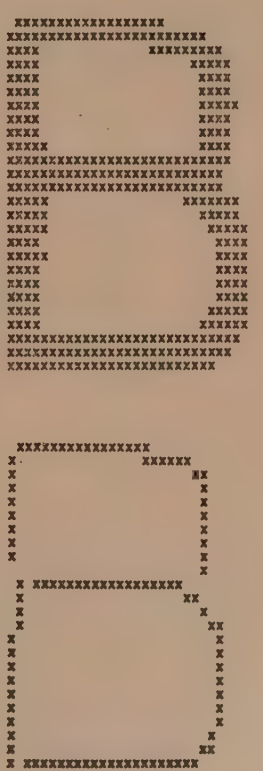


Fig. 20.—Reduction of a thick to a thin character.



Fig. 21.—Broken or noisy characters.

Whilst the method may be applied to the recognition of any pattern, there are certain restrictions in the present system. It can deal with the small letters of the English alphabet, the Greek alphabet and other similar sets of symbols, such as those shown in Fig. 22, but it fails with symbols having a large amount of detail. The system as arranged breaks down with complex patterns because the scanner has insufficient resolving power and there is not provision, in the present programme, for enough groups. Both these restrictions can be easily removed in future systems based on the same principles.

An important concept in the whole recognition system is the reduction of a two-dimensional pattern to a one-dimensional pattern. This reduction occurs in the scanning process in which the statements are prepared. The statements take the form of one-dimensional patterns, and it is these statements which are compared with those referring to the standard library patterns in the recognition section of the system. This is of twofold importance: first, the production of the statement is a process which is unaffected by the rotation of the pattern, or by variations in its size or precise form; secondly, the recognition problem is reduced to one of comparing one-dimensional patterns. This reduction makes the comparison possible in a reasonable time, which is particularly important when the library of standard patterns becomes large.

The whole operation is performed by a programme on a medium-speed digital computer. The programme has approximately 4000 instructions, and the time to recognize a pattern is of the order of 60 seconds, depending on the complexity of the pattern and the size of the library.

This time may be reduced by performing certain sections of the programme with special equipment. For example, the scanning of the pattern is a slow process on a digital computer and could be considerably speeded up. The time taken could be further reduced by performing the programme on higher-speed computers now being built. The reduction in time on such a machine would be by a factor of a thousandfold at least.

The present system can, in principle, be used for the recognition of handwriting. It would be necessary to extend the programme so that a greater number of groups could be handled, and provision would have to be made for dealing with complete words. With bad handwriting, interpretation is only possible, even for human readers, by using the context.

(5) REFERENCES

(1) Loeb, J.: 'Communication Theory of Transmission of Simple Drawings', in 'Communication Theory' (Butterworths Scientific Publications, 1953), p. 317.

(2) Discussion of above, p. 322.

(3) Shepard, D. H., Bargh, P. F., and Heasly, C. C.: 'A Reliable Character Sensing System for Documents Prepared on Conventional Business Devices', *Wescon Convention Record*, 1957, Part 4, p. 111.

(4) Chow, C. K.: 'Optimum Character Recognition System using Decision Functions', *ibid.*, p. 121.

(5) Dineen, G. P.: 'Programming Pattern Recognition', *Proceedings of the Western Joint Computer Conference*, Los Angeles, March, 1955.

(6) Bailey, C. E., and Norrie, G. O.: 'Automatic Reading of Typed or Printed Characters', British Institution of Radio Engineers Convention on Electronics in Automation, June, 1957.

(7) Greanias, E. C., Hoppel, C. J., Kloomok, M., and



Fig. 22

OSBORNE, J. S.: 'The Design of the Logic for the Recognition of Printed Characters by Simulation', *Proceedings I.E.E.*, Paper No. 2149 M, November, 1956 (**103** B, Suppl. 3, p. 456).

(8) KIRSCH, R. A., CAHN, L., RAY, C., and URBAN, G. H.: 'Experiments in Processing Pictorial Information with a Digital Computer', *Proceedings of the Eastern Joint Computer Conference*, 1957, p. 221.

(9) 'Magnetic Ink Character Recognition.' The Technical Sub-Committee on Mechanization of Check Handling of the American Bankers Association. *Computers and Automation*, 1956, **5**, p. 10.

(10) SELFRIDGE, O. G.: 'Pattern Recognition and Modern Computers', *Proceedings of the Western Joint Computer Conference*, Los Angeles, March, 1955.

(11) UTTLEY, A. M.: 'Temporal and Spatial Patterns in a Conditional Probability Machine', Automata Studies (Princeton University Press, 1956), p. 277.

621.142 : 621.382

The Institution of Electrical Engineers
Paper No. 2593 M
Apr. 1958

OPERATING EXPERIENCE WITH A TRANSISTOR DIGITAL COMPUTER

By R. C. M. BARNES, B.Sc.(Eng.), Associate Member, and J. H. STEPHEN, Ph.D., Graduate.

(*The paper was first received 4th December, 1957, and in revised form 22nd January, 1958. It was published in April, 1958, and was read before the* MEASUREMENT AND CONTROL SECTION *4th November, the* SOUTH-WEST SCOTLAND SUB-CENTRE *4th November, and the* SOUTH-EAST SCOTLAND SUB-CENTRE *5th November, 1958.*)

SUMMARY

A small digital computer using transistors as the only active circuit elements has been operating for a year. The paper describes the performance of this computer and, in particular, the failure rate of point-contact transistors. Earlier information on this computer is supplemented by details of circuit developments.

(1) INTRODUCTION

A small digital computer, Cadet, using transistors as the only active circuit elements, was described in earlier papers.[1,2,3] It is a serial, fixed-point, binary computer in which, apart from a few transistor registers in the control and arithmetic units, storage is entirely on one magnetic drum. This drum provides a working store of 410 words and a main store of 16 384 words. The typical arithmetic instruction specifies two addresses in the working store from which operands are to be read, an address to which the sum, difference or product of the operands is to be sent, and the address of the next instruction. Blocks of 64 words may be copied from the main store to the working store and vice versa.

Cadet is a laboratory model with an untidy, extended, 2-dimensional layout and long, unscreened interconnection wiring. It has been in regular use since August, 1956, operating mainly on one programme concerned with crystal structure analysis. It is seldom switched off and runs without attention during the night and at week-ends.

This paper reports the performance of Cadet over a period of a year, with analyses of serviceability and transistor failures. The circuit design has been changed in a few details, mainly to take advantage of later types of transistor, and some of these developments are described in Section 4. Various minor additions to the computer have increased the number of transistors to 324 point-contact and 76 junction type.

(2) RELIABILITY

The state of the computer since August, 1956, is summarized in Fig. 1. In this diagram non-productive serviceable time (*d*) includes periods when no work was available and also a few periods when the computer was 'idle but serviceable' owing to an unattended production run having been stopped by an isolated error. Isolated errors can cause a serious loss of productive time because Cadet is unattended for more than 75% of the total time. The improved serviceability since January, 1957, results mainly from various measures to combat isolated errors, principally by better preventive maintenance and by programme and engineering innovations to ensure that the computer can usually restart after being halted by an error. Time for checking new programmes has been included as non-productive, but accounts for only 2½% of the total time.

Unserviceable time is divided into (*b*) time when the computer is actually handed over to the engineers and (*c*) time when engineering attention is required. This latter category includes time wasted in unreliable computing or in a closed loop of instructions due to a component failure while unattended. Engineering time (*b*) includes a routine daily marginal test taking 30 minutes.

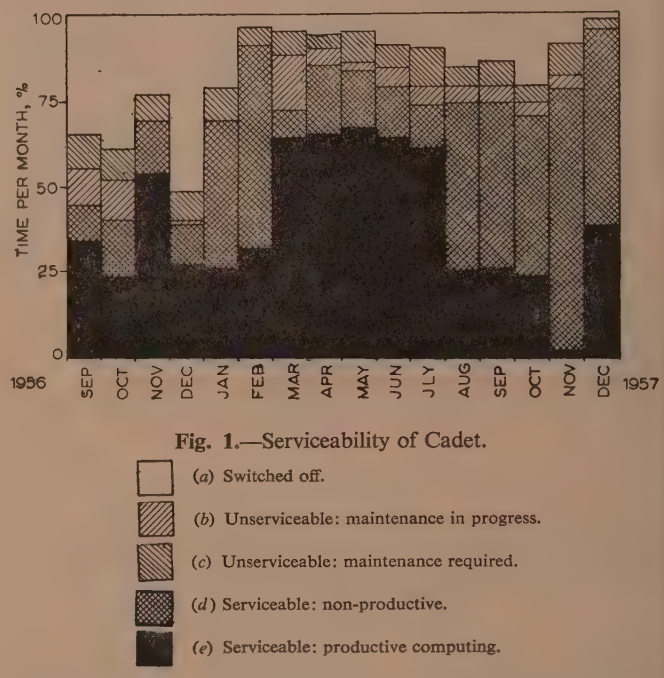


Fig. 1.—Serviceability of Cadet.

The computer was switched off during the remaining 5–10% of the time. A safety trip circuit disconnects the main d.c. power supply if the basic rhythm of reading and obeying instructions stops while the computer is unattended, and the operation of this safety circuit accounts for the majority of the 'off' time. The safety circuit does not switch off the driving motor of the magnetic drum, as a precaution against indentation of the heavily loaded ball bearings, nor does it switch off the power supply to the earlier stages of the reading amplifiers, thus avoiding the slight risk that signals on the clock tracks might be erased by current surges in the reading heads. If it is necessary to switch off the amplifier power supply during maintenance periods, the reading heads are first isolated from the amplifiers.

It will be seen from Fig. 1 that there was an initial period during which intermittent faults resulted in much 'off' time. These faults were traced, and since then Cadet has been switched on for more than 80% of the total time each month, and has been serviceable for more than 70% of each month, rising to 91% and 95% in two months.

(2.1) Component Failures

The most frequent component failures have been among point-contact transistors, where 41 transistors from the 324 point-contact type in the computer were replaced in 12 months.

Mr. Barnes and Dr. Stephen are at the U.K.A.E.A. Atomic Energy Research Establishment.

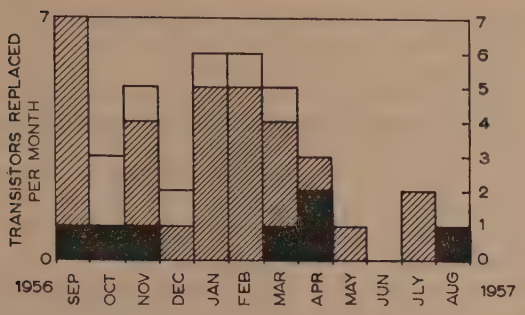


Fig. 2.—Monthly record of point-contact transistor replacements.

Accidentally destroyed.

Replaced during preventive maintenance.

Failed in service.

Transistors have only been replaced when their characteristics were outside the initial acceptance test limits. It will be seen from Fig. 2, which shows the point-contact transistor failures for each month, that only seven transistors failed while the computer was in use, but 27 were found during preventive maintenance to be outside specification. Seven transistors were destroyed by accidental short-circuits, which were formerly quite frequent. Now that engineering time is only a few hours per week, and precautions are taken to guard against short-circuits due to causes such as crocodile clips on oscillograph leads, these accidents are rare. An analysis of the 34 transistors which failed naturally is given in Section 3. During the same period, three junction transistors were replaced. The transistor failure rates, excluding accidental destruction, are therefore 1·2% per 1000 hours for point-contact types and 0·5% for a small sample of junction types.

No detailed record has been kept of other failures, but it is possible to make some qualitative comments. Failures of resistors and capacitors have been negligible. A few point-contact germanium diodes have been replaced, but in most cases it seems probable that the faulty diodes had been accidentally overloaded or that an unsuitable type had been used.

(2.2) Magnetic-Drum Store

Many failures of the computer have been due to the magnetic-drum store. The signal read from the drum has often been incorrectly phased with respect to the strobing pulse. This type of fault is due to mechanical movement of the blocks of heads or to waveforms which are far from the ideal shapes described in the earlier paper,[3] and can usually be attributed to the operating conditions of the drum store being much more stringent than those for which the mechanical design was intended.[4] In particular, 180 digits are recorded per inch on this drum, demanding stable head mountings and an unusually small clearance between the heads and the surface of the drum. In the working store it is necessary to be able to write isolated digits without disturbing digits in adjacent positions. Our early experiments used a drum[3] on which the heads were individually adjustable and could be allowed to touch the surface of the drum, giving the minimum clearance. With the present drum, which has slightly resilient head-mountings, a perceptible clearance must be maintained to avoid 'chattering', and this reduces the safety margin when writing isolated digits at the close digit packing. As 17 heads are assembled into a block and are not individually adjustable, the clearances between the heads and the drum are not uniform, and the performance on some tracks is barely acceptable. Programmes designed to test the ability to write isolated digits have shown that the waveform at some reading heads deteriorates progressively. Care must therefore be taken, in any normal programme which requires writing into isolated digit positions, to ensure occasional rewriting of all digits to standardize the waveforms. Not all working store tracks suffer from this defect, and it is probable that reliable operation could be obtained by minor changes to the mechanical design, allowing either closer and more consistent clearances between heads and drum surface, or a reduction in the number of digits per inch.

Although the drum store has been responsible for many computer faults, it has, nevertheless, given long periods of trouble-free service in which any small changes of adjustments could be corrected during preventive maintenance. A simple thermostatically controlled enclosure protects the drum from normal variations of room temperature, but on several occasions this protection has been inadequate under extreme conditions. The resulting differential thermal expansion seriously disturbed the mechanical adjustments of the drum heads and a period of unreliable performance followed.

At first, silver relay contacts used for switching the low-level signals from the reading heads of the main store gave many high-resistance faults. Platinum contacts are now giving satisfactory service, and a diode switching scheme which was investigated for this purpose has not been completed. Some trouble was also experienced with interference due to sparking at relay contacts which switch the highly inductive magnets of the head-shifting mechanism, and a power transistor circuit, described in Section 4, was installed.

(2.3) Input-Output Equipment

The input-output punched-card equipment was deliberately kept as simple as possible. An end-fed card punch used as output organ, and a card verifier used as input, required only minor modifications to their internal wiring. Most of the connections between the computer and the punched-card machines are made through multi-way sockets which originally served the manual keyboards. Several features of these machines which give trouble while the computer is unattended would be acceptable in their normal applications, where an operator is always present and where cards pass only once through each machine. For instance, the card hoppers hold only 250 cards, and the card-feed mechanism does not always tolerate slightly worn programme cards. By making full use of the large backing store, it is usually possible to concentrate input-output operations into the periods when the computer is attended, and thus avoid these difficulties.

(2.4) Preventive Maintenance

The safety margins of most transistor circuits are checked by means of a marginal test which applies 50 c/s modulation to the more critical d.c. supply lines. This test can be switched into individual sections of the computer, but is more often applied to the whole of the computer in conjunction with test programmes, using random patterns of digits whenever possible. The marginal test has proved to be a valuable aid in bringing all circuits to a uniform standard of reliability.

The amplitude of the alternating marginal test voltage can be varied, and is now set during the daily test period at 85% of the voltage at which general failure of circuits throughout the computer can be expected. Under these test conditions the +12-volt supply varies by ±2·0 volts, and the +1·5-volt supply by ±0·36 volt, to give two examples. An error-free run of at least 30 min under daily test conditions is the minimum acceptable performance.

At irregular intervals of approximately two months an inspection of the collector waveforms of all transistor circuits has detected a few transistors with characteristics outside the acceptance limits and has resulted in a noticeable improvement in reliability. The effect of external electrical interference has become less serious as unreliable transistors and design weaknesses have been detected by marginal testing and corrected. The interference from relay circuits in the punched-card machines was suppressed with considerable difficulty. It is still necessary to avoid switching inductive loads and fluorescent-lamp starting circuits near the computer, but otherwise normal laboratory activities are permitted.

(3) ANALYSIS AND DISCUSSION OF TRANSISTOR FAILURES

(3.1) Point-Contact Transistors

The 34 transistors which failed in Cadet between September, 1956, and August, 1957 (excluding those damaged by accidents), can be classified as in Table 1.

Table 1

TRANSISTOR FAILURES

Reason	Type TP1	Type EW61
Poor collector bottoming potential	7	5
High collector-to-base leakage current I_{c0}		2
Inside test limits but failed to trigger on	1	
Slow turn-on time	1	
Slow turn-off time	1	
'Floating emitter effect'	17	
Total failures	27	7
Population in computer	224	100
Failures as percentage of population	12%	7%

The poor collector bottoming potential is caused by a reduction in the current gain of the transistor at low collector voltages. The computer circuits require the collector bottoming potential to be less than 2 volts for an emitter current of 3 mA and a collector current of 5 mA. Any transistor failing this test is rejected. The bistable circuits have been designed to use transistors in which the collector-to-base leakage current, I_{c0}, is less than 2 mA for a collector-to-base voltage of 18 volts. If I_{c0} exceeds 2 mA the base bias circuit may fail to hold the transistor cut-off. The transistor which failed to trigger satisfactorily had an abnormally low value of I_{c0}, a characteristic which was noted earlier with other transistors. The transistor with slow turn-on time was removed during maintenance as it could possibly cause trouble. The transistor with slow turn-off time was in the pulse generator and was replaced although it had not caused the computer to fail.

The most serious cause of failure has been the 'floating emitter effect' which has been described in the literature by Ryder and Sittner.[5] They name it 'sleeping sickness' and discuss its occurrence in grown *n–p–n* transistors. So far this effect has been observed in Cadet in the point-contact transistors from one manufacturer only. A description of the effect is as follows.

A normal transistor with its base earthed and emitter open-circuited has a negative emitter voltage of between 100 and 200 mV when the normal working collector voltage (−18 volts) is applied. If the transistor has the 'floating emitter' characteristic, the emitter voltage slowly becomes more negative and may be several volts negative after the collector voltage has been applied for a minute or more. The emitter potential remains negative if a small forward emitter current is applied, but, as this current is increased, the emitter eventually becomes positive with respect to the base. Measurements of current gain[6] made while the emitter is negative show that the gain can be as high as ten, dropping to between two and three as the emitter current is increased to 1 mA. The existence of current gain greater than unity when the emitter is negative means that the effect is not simply due to a high-resistance leak between emitter and collector. The cause of this effect is not known, but one possible explanation is that it is caused by current leakage from the emitter along a surface layer on the germanium, and the occurrence of carrier injection from this layer into the germanium near the collector. Such leakage current would reduce the apparent contact spacing and thus increase the effective base resistance, causing instability.

When the bistable circuit in the computer is not conducting, the base of the transistor is biased 1·5 volts positive with respect to the anode of the emitter series diode. In the monostable circuit this bias is 2·25 volts. If the emitter potential of a transistor in these circuits tends to float negative with respect to the base by more than these voltages, the emitter series diode will conduct and emitter current will flow. Since the current gain of the transistor exceeds unity the circuit will switch on. In some cases spurious switching has occurred when the floating emitter voltage was less than the bias voltage, due to the presence of small positive voltage spikes at the emitter. These 'spikes' are due to leakage of pulses through diode gates or to current surges in power supply lines. Because the floating emitter effect builds up slowly when the collector potential is applied, it tends to produce computer failures which are very sensitive to the sequence of digits in the faulty stage.

(3.2) Junction Transistors

Three junction transistors were removed from the computer during the same period in which 41 point-transistors failed. All three transistors showed the same fault, a very considerable increase in the collector-to-base leakage current. The performance of the junction transistors has otherwise been very satisfactory.

(4) CIRCUIT DEVELOPMENTS

(4.1) Clock-Pulse Frequency

The bistable and monostable circuits in the computer are unchanged from the circuits described in an earlier paper.[2] One major change in the design of the computer has been to reduce the drum speed from 1 450 to 950 r.p.m., which reduced the basic clock frequency from 58 to 40 kc/s. This change was made to ease the decoding of the information from the drum. The 'on' and 'off' times of the amplitude-limited waveforms were thus lengthened, but, as the original strobe width of 2 microsec was retained, the margin of safety for decoding was improved.[3]

(4.2) Reading Amplifiers

The reading amplifiers used two junction transistors and two point-contact transistors.[3] Normally the gain of these amplifiers was adequate, but in time the characteristics of the point-contact transistors changed sufficiently to reduce the gain or produce instability.

A junction-transistor amplifier (Fig. 3) was designed to replace the original amplifier. Five h.f. junction transistors are used, four in common-emitter connection with current feedback over the first and second pairs, and one as an emitter-follower output stage. D.C. feedback is applied over the whole amplifier to stabilize the operating points. The reading head is matched

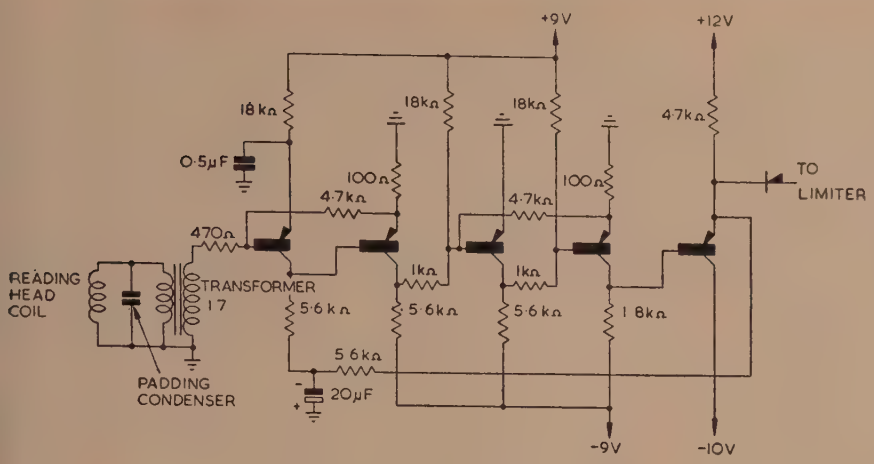


Fig. 3.—Junction-transistor reading amplifier.

to the amplifier by a 1/7 transformer. The 470-ohm resistor and the input impedance of the amplifier (about 100 ohms) provide the necessary damping for the reading head. The overall current gain is 11 000 and the h.f. cut-off is 200 kc/s. Over the frequency range of the read signals (20–40 kc/s) the amplifier response is constant. As the amplifier is d.c.-coupled and uses only an input transformer, it overloads by clipping the peaks of the waveform and not distorting it in any other way. One of these amplifiers has been operating satisfactorily for nearly a year.

As the replacement of the old amplifiers by the type just described would have entailed building fresh equipment, an effort was made to adapt the old amplifiers to use junction transistors. Early experiments with the amplifiers showed that the junction transistors then available had large base resistances and were unsuitable in place of the point-contact transistors. Since then, junction transistors with base resistances of about 100 ohms have become available, and these satisfactorily replace the two point-contact transistors, except that the transistor current gain per stage drops from 1·5 to less than 1·0. To overcome this loss in gain the junction transistor in the first stage is replaced by an h.f. junction transistor in the common-emitter connection with current feedback (Fig. 4). The gain of this stage is roughly two, and this can be increased by increasing R_1, a facility which has made it possible to use working store tracks with very low outputs. The input impedance is about 50 ohms, so an extra 470-ohm resistor is added to provide adequate damping for the reading head. This modification, which is simple to make, is proving very satisfactory, and it is being incorporated in all the original amplifiers.

(4.3) Main-Store Track-Selection Circuits

The main store has 256 tracks on the drum, each storing 64 words.[3] A 9-bit transistor register is used to select one of these tracks for writing or reading. The main store head block contains 16 heads which are used for both writing and reading. The block can be moved parallel to the axis of the drum by an electromagnetic head-shifting mechanism[4] to 16 discrete levels, and thus each head has 16 tracks associated with it.

A 'tree' of 3000-type relays controlled by four bits of the register selects the head to be used. The next four bits control, through 3000-type relays, the head shifter which moves the block of heads to the selected level, and the remaining bit in the register switches the circuits for writing or reading.

Fig. 5 shows a stage of the 9-bit transistor register which operates a 3000-type relay. A standard point-contact-transistor bistable circuit is used to provide a 7 mA base current for a junction transistor in the common-emitter connection with the relay coil in its collector circuit. The junction transistor saturates when switched on, so that the full supply voltage of 27 volts is applied across the relay coil to make it operate rapidly. As the electrolytic capacitor recharges, the current in the coil decreases to a holding current of 30 mA. The 3·9-kilohm resistor and the diode to earth in the junction-transistor base circuit ensure that this transistor is biased off when the point-contact transistor is not conducting.

A circuit using power transistors (Fig. 6), similar to the one just described, has been designed to energize the magnet coils of the head shifter. Presently available power transistors are limited by a maximum collector-to-base voltage of 30 volts. As the head shifter operates with a 60-volt supply, the circuit of Fig. 6 therefore uses two transistors in series. A potential divider provides a 30-volt tapping for the base of the second transistor. Both transistors have base currents of 17 mA when conducting, to ensure that they are bottomed. One power transistor could be used if the head shifter were dismantled and rewound for 30 volts. An extra junction transistor would then be required to provide the larger base current for the power transistor.

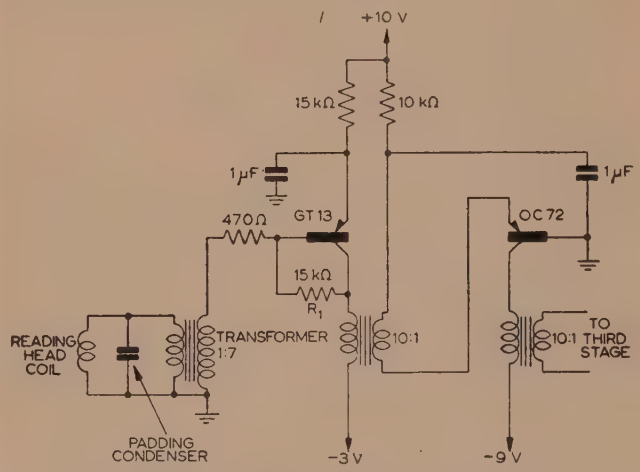


Fig. 4.—Circuit modifications to the first stage of the existing amplifier.

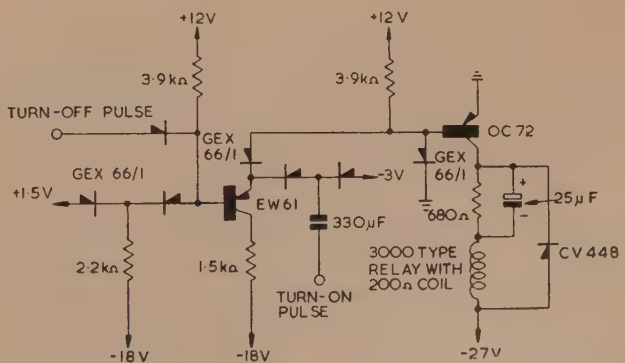


Fig. 5.—Bistable circuit for operating a 3000-type relay.

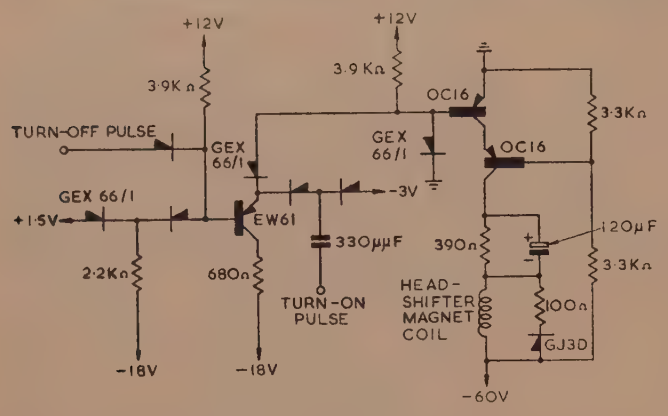


Fig. 6.—Bistable circuit for energizing a coil of the head shifter.

An important mechanical modification to the head shifter was the fitting of a p.t.f.e. liner to the tube in which the cylindrical armatures slide. The liner produced a considerable improvement in the speed and general operation of the head shifter.

(4.4) Junction-Transistor Bistable Circuit for Replacing a Point-Contact Transistor

A unit including one *p–n–p* and one *n–p–n* junction transistor has been developed as a possible replacement for a point-contact transistor in the bistable circuit of the computer. The unit was designed on the following basis:

(*a*) Its use should involve the minimum of change to the existing bistable circuit.
(*b*) The switching speed and sensitivity should be the same as those of the point-contact transistor.
(*c*) It should use the existing 18-volt collector supply.
(*d*) It should pass 30 mA and saturate to not more than 3 volts at this current.

High-frequency transistors must be used to obtain the same order of switching speed as is obtained with point-contact transistors. While the *n–p–n* transistors are rated at 20 volts collector-to-base, the *p–n–p* h.f. transistors are rated at only 15 volts, and it is therefore necessary to select suitable transistors by measuring collector-to-base leakage current and collector-to-emitter 'punch through' voltage.

It is intended that the substitution units will be used as replacements for faulty point-contact transistors. There is little to be gained by replacing all point-contact transistors in bistable circuits by junction transistor units, since the failure rate of point-contact transistors in all circuits (including monostable circuits and amplifiers) was only seven transistors in a year. At present three substitution units have been giving satisfactory service in the computer for about six months.

(5) UTILIZATION

In this Section some of the unusual programming features of Cadet are reviewed, and an outline of the programme which occupies most of the running time leads to a description of measures adopted to make the best use of unattended operation.

(5.1) Working-Store Addresses

Cadet has a 3 + 1 address instruction which specifies the addresses of two operands, the address of a destination for the result of an arithmetic or logic operation and, with a restricted choice, the address of the next instruction. The first feature which will be discussed is the addressing of the operands and destination.

The working store consists of eight tracks on the magnetic drum. Each track is divided into eight sectors, and in each sector there are eight interleaved words, referred to as eight 'phases'. The track, sector and phase of an address in the working store are therefore specified by three octal digits.

Two tracks have special features. On one track, the single regenerative loop (s.r.l.), the same eight words appear in each sector and are therefore always available to participate in operations. On another track, the double regenerative loop (d.r.l.), the programmer is usually only concerned with one particular phase, in which a double-length word is repeated in four pairs of sectors. This phase is used as the product accumulator during multiplication. While it might be possible to write a programme in which the majority of addresses refer to the six normal tracks of the working store, with only occasional use of the s.r.l. when its immediate access feature is essential, it is certain that such a programme would be inelegant and difficult to construct. Programming is greatly simplified by placing the eight most frequently used quantities in the s.r.l. A typical programme then contains many references to the s.r.l., references to the d.r.l. which are mainly concerned with multiplication, and a minority of references to other tracks. This situation will also apply to many other computers with multi-address instructions and storage which is not uniformly or randomly accessible. It would not greatly change the nature of the programmes for Cadet if one operand were always read from the s.r.l. track. The principal effect would be to require duplication of some operation codes in which the operands are not interchangeable, e.g. the subtraction operation A − B → D would require supplementing with −A + B → D if B always referred to the s.r.l. track. The operand address part of the instruction would then require one less octal digit and it would be possible to expand the next-instruction address and avoid difficulties which are described below.

(5.2) Address of Next Instruction

As the store is not randomly accessible, it is necessary to place numbers and instructions so that they become available with minimum waiting time. All operations do not take the same time to perform, so that successive instructions cannot be drawn from addresses in a regular sequence. Each instruction must therefore state either the relative or the absolute address of the next, but the normal instruction word has only one octal digit, the 'increment digit', available for this purpose. Successive instructions are normally read from the same track, and the increment digit allows a choice from only eight of the 63 possible addresses for the next instruction. The requirements of 'optimum programming' are best met by having the increment digit select one of the eight most useful relative addresses. The system suffers from two weaknesses. The restricted choice is sometimes inadequate, for example when non-optimum timing is required, and the increment digit is often incorrectly coded because the relationship between the increment codes and the addresses which they select is not easily recognized.

The first weakness has been partially compensated by jump instructions which specify a full next-instruction address, and also by some instructions which require only one operand address, so that two octal digits are available to state the sector and phase of the next-instruction address. If two octal digits were available, so that the sector and phase of the next-instruction address could be stated in each instruction, it would be much easier to obtain optimum timing and efficient utilization of a greater proportion of each track, and would also allow more freedom when modifying programmes.

(5.3) Writing Programmes for Use in any Track

It is convenient to copy from the main store to the working store the contents of a whole track containing instructions and

some numbers closely associated with them. A widely used subroutine may not always occupy the same working store track, and the programmer cannot, therefore, know in advance exactly how to code the track digits of the addresses of operands which accompany the subroutine. Recoding or programmed modification of such track digits to suit the appropriate working store track has been avoided by a convention which marks these variable track digits with a special code. Track digit 0 has been given this special significance, and very few extra circuit components were needed to detect zero in the relevant track digits of the P.I. register and to substitute the appropriate track digit, which is that contained in the C.I. register. This very useful feature was achieved without sacrificing the use of Track 0 of the working store, although its use is now slightly restricted, as it cannot be used as the operand or destination address of most instructions written on other tracks. This restriction has little effect on programme design.

(5.4) Choice of Work for Unattended Operation

The error rate of Cadet can be kept consistently low by careful preventive maintenance, and the fire risk from circuits operating at low power and low voltage is negligible. Long unattended runs are therefore a practical proposition and enable the output of work to be trebled with no increase in operating cost. The input-output equipment is not well suited for unattended operation, particularly on account of its small card capacity, and this has led to a choice of work in which the capacity of the backing store is adequate to hold programme, initial data, and results pertaining to a run of 60 hours or more.

An application which provides an almost continuous load for Cadet, and which satisfies the requirements for unattended operation, is the processing of neutron diffraction data by a Fourier synthesis method[7] in order to obtain the distribution of atomic nuclei in single crystals. The main programme loop, computing an expression of the form $\sum_{n=1}^{n=180} A_n \cos 2\pi(h_n x + k_n y)$ is comparatively simple. The initial data, A_n, h_n, k_n, occupy less than 1/32 of the backing store. Repeating this summation for 61 values of x and 31 values of y, 1 891 cases in all, requires a running time of approximately 80 hours and uses less than a quarter of the backing store to hold all the information which is transferred to punched cards as a separate operation.

Cadet has operated on this programme for most of the period during which a detailed log has been kept, and it has been regular practice to complete the majority of one run during the weekend, running from Friday evening to Monday morning without supervision by engineering or operating personnel. Some of the practical difficulties encountered in this mode of operation, and the remedies adopted, will now be described.

(5.5) Correction of Control Errors

One obvious difficulty is that an isolated control error may break the sequence of instructions. There tends to be an appreciable number of unused addresses in the working store, partly because it is convenient to store one (or more) reasonably self-contained sections of programme on each track without attempting to use all the addresses, and also because the allocation of addresses to suit optimum programming frequently prevents efficient utilization of the store. Consequently, a control error is quite likely to result in an unused address being selected as the source of the next instruction. It is therefore usual to ensure that unused addresses contain zero rather than irrelevant words left over from some previous programme. A few extra diode gates were added to detect all redundant operation codes, including 00 and 77 which were deliberately chosen as redundant codes since they are the most probable results of many control errors. The detection of a redundant operation code can either be allowed to stop the computer, which is the normal practice while staff are present to investigate the cause of the error, or it can cause the computer to revert to the initial input instructions, so that the action to be taken after a control error can be indicated by cards left in the card-reader.

(5.6) Numerical Errors

These are less easily detected, but they do not, in general, interrupt the computing sequence so drastically as do control errors. In the case of the crystal structure analysis for which this computer has mainly been used, the general form of the results can be predicted, so that a graphical plot of the results (using an automatic curve plotter with punched-card input) will immediately reveal any gross errors. As a further safeguard a programmed check is incorporated which repeats some sections of the computation until the same result is obtained twice in succession. A serious arithmetic breakdown usually results in the computer being unable to pass the next programmed check, and this has been found in practice to guard against the production of erroneous results. For this particular problem a partial check is adequate, since, by increasing confidence in alternate points of the plotted results, it makes it unlikely that an error in the unchecked points would pass unnoticed.

The adder can add 16 words concurrently (eight interleaved words from each of two tracks), and this feature is used in multiplication and also in an operation which sums all 64 words on one track. This special addition makes it very easy to form a check sum of the contents of a track, and is extensively used as a means of detecting any errors introduced when trackfuls of data or results are copied into, or out of, the main store.

(5.7) Reporting Information about Failures

As unattended operation of Cadet greatly exceeds attended operation, it is often difficult to collect sufficient evidence about intermittent fault conditions. It was soon found, for instance, that when the computer was shut down by the automatic alarm circuit there was no indication of the instruction which led to the failure, or of the fact that the complete failure might have been preceded by a period of faulty computing. Two minor additions were therefore made to the computer to record information about faults. In the d.r.l. track two spare phases in each sector are used to record the last instruction read in that sector and the previous contents of the instruction register. When a fault occurs, the record in these two phases shows the last instruction and also seven other fairly recent instructions, although these are not necessarily consecutive. From this evidence it is usually possible to deduce the sequence of events immediately preceding the fault. This feature was introduced as an aid to maintenance, but it is also helpful when testing new programmes, particularly by reducing the number of occasions on which it is necessary to re-run a long programme in order to obtain information about instructions which precede a programme error.

Programmes which have been used on Cadet have so far been of the type in which the main store tracks are selected in a regular sequence. A record of this sequence gives useful information about the state of the computer. The main store-track selection system is controlled by a transistor register (see Section 4.3), and an analogue indication of the contents of part of this register is obtained by summing weighted currents and is recorded on a recording milliammeter. At a chart speed of 1 in/h this shows a repetitive pattern which is characteristic of the programme and is, in effect, an automatic log of the use of the computer. Irregularities in this record of main store track selection are often the only evidence of occasional errors which

have not been serious enough to stop the computer, and are a warning that extra preventive maintenance may be necessary.

(6) CONCLUSIONS

One objective when designing Cadet was to produce a computer in which reliability and simplicity were of more importance than high computing speed, and which would be capable of running for long periods without attention and with very little maintenance. The performance of Cadet, as reported in this paper, shows that this has been achieved. Cadet was also intended to provide experience in the use of transistors, and data on their reliability. The failure rate of point-contact transistors has been higher than was expected, but is of the same order as that which has been quoted[8] for standard commercial thermionic valves in digital computers. Junction transistors, diodes and other components have proved to be extremely reliable in circuits operating at low voltages and low power dissipation. Experience with the magnetic-drum store has shown that the reliability is barely adequate in the working store, where the interleaved digit system makes it necessary to record isolated digits without disturbing those in adjacent positions. This has proved to be difficult on a drum which has 180 digits per inch and on which the reading and writing heads are mounted so that neither their circumferential spacing nor their clearance from the drum surface can be adjusted, and held, to sufficiently fine limits. High digit density alone, as in the main store, has not led to serious difficulties. The transistor circuits associated with the drum store have, like the logic circuits, given a satisfactory performance.

(7) REFERENCES

(1) Cooke-Yarborough, E. H., Barnes, R. C. M., Stephen, J., and Howells, G. A.: 'A Transistor Digital Computer', *Proceedings I.E.E.*, Paper No. 1964 M, January, 1956 (**103**, Part B, Supplement No. 3, p. 364).

(2) Barnes, R. C. M., Howells, G. A., and Cooke-Yarborough, E. H.: 'Transistor Arithmetic Circuits for an Interleaved-Digit Computer', *ibid.*, Paper No. 1992 M, February, 1956 (**103**, Part B, Supplement No. 3, p. 371).

(3) Stephen, J. H., and Cooke-Yarborough, E. H.: 'An Interleaved-Digit Magnetic Drum Store for a Transistor Digital Computer', *ibid.*, Paper No. 1999 M, February, 1956 (**103**, Part B, Supplement No. 3, p. 382).

(4) Clayden, D. O., Page, L. J., and Osborne, C. F.: 'The Magnetic Storage Drum on the Ace Pilot Model', *ibid.*, Paper No. 2090 M, March, 1956 (**103**, Part B, Supplement No. 3, p. 509).

(5) Ryder, R. M., and Sittner, W. R.: 'Transistor Reliability Studies', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 414.

(6) Cooke-Yarborough, E. H., Florida, C. D., and Stephen, J. H.: 'The Measurement of the Small-Signal Characteristics of Transistors', *Proceedings I.E.E.*, Paper No. 1614 R, September, 1954 (**101**, Part III, p. 288).

(7) Bacon, G. E., and Pease, R. S.: 'Neutron Diffraction Study by Fourier Synthesis', *Proceedings of the Royal Society*, A, 1953, **220**, p. 397.

(8) Knight, L.: 'Valve Reliability in Digital Calculating Machines', *Electronic Engineering*, 1954, **26**, p. 9.

[The discussion on the above paper will be found on page 237.]

621.374.32 : 681.142

The Institution of Electrical Engineers
Paper No. 2672 M
June 1958
©

A NEW HIGH-SPEED DIGITAL TECHNIQUE FOR COMPUTER USE

By D. ELDRIDGE, B.A.

(*The paper was first received* 14*th February, and in revised form* 2*nd April*, 1958. *It was published in June*, 1958, *and was read before the* MEASUREMENT AND CONTROL SECTION 4*th November*, 1958.)

SUMMARY

A new method is described for realizing logical functions using square-loop ferrite cores and transistors. Complex circuits can be built up from identical elements, each consisting of one core, one transistor, three diodes and one resistor. The technique differs from earlier methods in that the current required to set a core to the '1' state is derived from a common supply and not from a previous stage.

The control of digit transfer between stages is effected by a transistor whose collector and emitter are connected across one winding of the core. When a '1' is to be set in the core, the transistor is arranged to appear as an open-circuit, in which condition flux reversal is possible; when a '0' is to be set, the transistor appears as a short-circuit preventing reversal of flux.

The core is reset to '0' at a controlled rate, producing a standardized voltage across the output windings. Logical operations are carried out by the analogue addition of these voltages. The system requires only two low-voltage d.c. supplies and is tolerant of voltage and component variations. It operates reliably at digit rates of 500 kc/s. The paper also describes an experimental application of the method, designed to test its performance.

LIST OF SYMBOLS

Φ_s = Flux swing of core from positive remanence to negative saturation, maxwells.
F_0 = Magnetomotive force required to drive the core at a restricted rate, AT.
F = M.M.F. required to reverse the core state completely in time t, AT.
$\hat{F}_1$ = Maximum permissible set m.m.f. to core, AT.
F_1 = Minimum permissible set m.m.f. to core, AT.
F_{in} = Set m.m.f. to core, AT.
F_{out} = Reset m.m.f. to core, AT.
t = Set pulse length, microsec.
T = Reset pulse length, microsec.
v_{in} = Input amplitude that allows a stage to be set to '1' condition, volts.
v = Limit to voltage across control winding during reset period, volts.
v_0 = Voltage developed across a saturated transistor and series diode, volts.
y_a = Permissible number of outputs of type (a).
y_b = Permissible number of outputs of type (b).
y_c = Permissible number of outputs of type (c).
y_d = Permissible number of outputs of type (d).
I_b = Biasing current to base of transistor, amp.
$\hat{I}_c$ = Peak collector current of transistor when saturated, amp.
β_{max} = Maximum collector-to-base current gain allowed under pulse conditions.
C = Number of control-winding turns.
n = Number of output-winding turns.
f_{co} = Short-circuit common-base cut-off frequency, Mc/s.
f'_{co} = Short circuit common-emitter cut-off frequency, Mc/s.

(1) INTRODUCTION

It has long been realized that the design, construction and maintenance of digital computers can be greatly simplified by restricting the number of different logical elements or 'building bricks' that are used. An additional advantage is obtained if the use of such units can be specified by simple mathematical rules without reference to the detailed circuits involved. In this way, the tasks of the logical designer and the circuit designer can be separated.

If elementary units are to be used successfully as logical links it is essential that the operation of each unit shall be independent of those to which it is connected. In addition, units should give sufficient output to drive many others simultaneously, and should be compatible, so as to permit their random interconnection.

To approach an optimum system as closely as possible, the number of components used in each element should be minimized. In particular, standard units should not contain components that are only rarely used. To obtain a high operating speed, especially in systems handling information in a serial mode, a high digit rate is necessary. However, for parallel operation, the time required to perform the common logical functions may be more significant. It is obviously desirable that power consumption be kept low and the effects due to parameter variation avoided. The technique to be described here is an attempt to meet these somewhat conflicting requirements.

As in many previously described circuits,[1,2,3,4] use is made of a magnetic material having a substantially square hysteresis loop. The outstanding feature of such material for this application is its possession of two stable states of magnetization close to the saturated states. Thus a binary digit may be represented by calling either state a '1' and the other state a '0'.

Many uses of a transistor in combination with a core have also been described.[5,6,7] Previously, however, the transistor has been used as an output device to provide power gain for driving a number of other similar stages. In the present technique it is used essentially as a switch, to which purpose it is particularly well suited, having a very low resistance in the conducting or 'on' state, and a very high resistance when biased to the non-conducting or 'off' state. Power gain is also achieved in this mode of operation. Stage output is provided by the voltage appearing across a winding on the core during flux reversal.

(2) BASIC PRINCIPLES

The circuit description which follows will be more readily understood from a consideration of the three basic principles used.

(a) A transistor can act as a switch between collector and emitter, which, when connected across a winding on a core, will provide an effective short-circuit while the transistor is conducting. Ideally, this would prevent any change of flux in the core, but, in practice, owing to the finite saturation resistance of the transistor, a very limited rate of change of flux is possible. When non-conducting, the transistor will present a high resis-

Mr. Eldridge, who was formerly with Ferranti Ltd., is now with I.B.M. British Laboratories.

tance across the core winding and change of flux in the core is possible.

(*b*) Control of a transistor so connected, and thereby of a core, may be effected by a potential applied to the base terminal. When this is negative with respect to the emitter potential (for a *p–n–p* transistor), the transistor conducts, and when positive, the transistor is non-conducting.

(*c*) With the 'switch' open the core may be used to give standardized voltage outputs across one or more output windings by the provision of a suitable voltage clamp on one winding. The outputs from various stages may be connected serially, either in phase or in opposition, to provide an analogue summation of the individual outputs. This sum, representing a particular logical function, may be used either to open or close the switch of the following stage.

(2.1) Basic Circuit and Operation for Single-Digit Transfer

Fig. 1 shows the basic circuit and Fig. 2 the associated current-drive waveforms. Typical input and output voltages are indicated. Set waveform *a* is in such a direction as to drive the core to one of its saturated conditions which will be called the

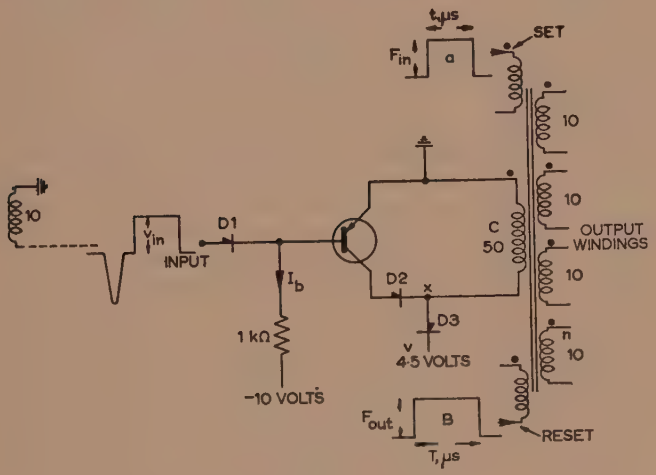


Fig. 1.—Basic circuit.

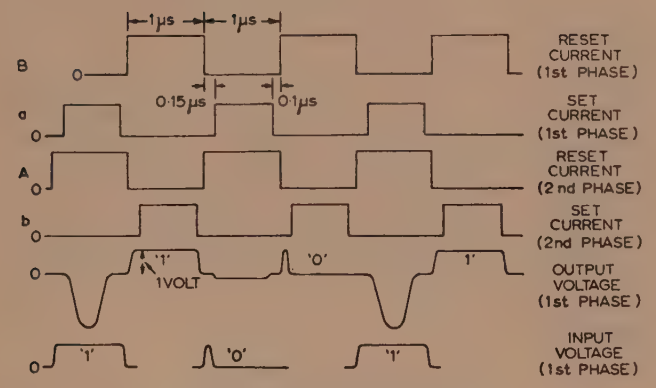


Fig. 2.—Timing waveforms.

'1' state. Reset waveform *B* is oppositely directed, driving the core to the other saturated condition, which will be called the '0' state. These two waveforms follow one another consecutively in time, with no overlap. Consider that initially the core is in the '0' state and that a direct voltage is applied at the input.

A positive input voltage raises the base potential of the transistor above the emitter potential and prevents any current from emitter to collector. Under these conditions, the impedance appearing across the control winding C is large so that its effect on the core may be ignored. As a result, the state of the core is continuously reversed by the driving pulses, set waveform *a* driving the core to the '1' state, and reset waveform *B* driving the core back to the '0' state.

In the absence of an input, the transistor is conducting and provides a short-circuit between its collector and emitter terminals. Thus diode D2 is effectively connected across the control winding C. The polarity of this diode is such as to prevent the potential at point X from falling below earth. As a result, the core cannot be set to the '1' state by set waveform *a*, which is in such a direction as to drive point X negative with respect to earth. In practice, owing to the finite voltage appearing across the transistor and diode D2, the core will be partially set. This effect is not cumulative, since the core is reset to a standard '0' condition by reset waveform *B*. Resetting is not affected by the state of the transistor, since diode D2 has a high reverse impedance.

The output voltages corresponding to the two states of the circuit described above are illustrated in Fig. 2. The '1' output voltage occurs when the transistor is rendered non-conducting for the duration of set waveform *a*, and consists of a negative-going pulse with a non-linear shape, followed by a positive-going pulse of well-defined amplitude. The negative output is due to the reversal of the core state from the '0' to the '1' condition by set waveform *a*. The subsequent positive output occurs when the core is reset to '0' by reset waveform *B*. Since a similar amount of flux is reversed in the core in each direction, the areas enclosed by the two outputs are approximately equal. However, the amplitude of the positive voltage is almost constant, since it is limited by the voltage-clamping diode D3.

The '0' output voltage occurs when the transistor is maintained in a conducting state for the duration of set waveform *a*. Under these conditions, the transistor acts as a short-circuit, and only a small amount of flux is set in the core. This partial setting is accompanied by a small negative output. In the following period the core is reset to a standard '0' condition producing a short positive output.

The binary state of the circuit described above may be transferred to another stage driven by a second pair of waveforms, *b* and *A*. For this purpose, the output of the first stage is connected directly to the input of the second. The initial part of the output from the first stage is negative, and does not affect the second stage owing to the high reverse impedance of the input diode D1. The subsequent positive output will be short if the first stage contains a '0', and long if it contains a '1'. To set the second stage to '1', the input must be held positive for the duration of set waveform *b*. Only a '1' input satisfies this condition (a '0' input finishes before the commencement of set waveform *b*). Thus a '1' input causes a '1' to be set in the second stage, while a '0' input leads only to partial setting. Reset waveform *A* drives the core back to the '0' state, producing output suitable for a further stage driven by waveforms *a* and *B*. By connecting the output of the second stage back to the input of the first stage, a staticizor is formed which circulates the binary state '1' or '0' continuously.

As already mentioned, D2 serves to isolate the transistor while the core is being reset. This is necessary, since a junction transistor is essentially a symmetrical device, able to pass current between emitter and collector in either direction. Thus the omission of D2 would load the core unnecessarily during the reset period and, more important, would lead to unwanted power dissipation in the transistor.

An important feature of the circuit is that the input pulse is not directly responsible for the reversal of the core state, but rather controls it by determining whether or not such reversal

can occur when the setting pulse is applied from an external supply. Thus, input to the unit is effectively strobed in time by the set waveform. Input at any other time has no effect on the core. This property is of the greatest importance when the input consists, as described below, of a composite waveform produced by the addition of several output voltages, each varying slightly in width and rise time. Also, of course, it is the means by which the finite '0' level is ignored.

In the absence of a positive input the transistor is conducting, owing to the biasing resistor connected to the base. In order that a core may be set, a '1' input must raise the base potential above earth and supply the bias current for the whole of the setting period (*a* or *b*). In addition, current must be supplied initially to the base of the transistor, but, even so, the required input current is low and does not lead to serious loading of the stage providing the input. Thus the driving m.m.f.'s need be little more than is necessary to reverse the state of an unloaded core. However, when many inputs are supplied from one unit, loading becomes more significant. This is considered in detail in Section 2.5.

(2.2) Logical Interconnection of Units

As explained above, each unit can give two outputs, corresponding to the binary digits '1' and '0'. However, a third possibility exists. By connecting an output winding in anti-phase, inverse outputs occur which may be labelled $-$'1' and $-$'0'. Logically, and in practice, the two zero states cannot be distinguished and may be taken to be identical. The third output, corresponding to $-$'1', provides a negative-going input voltage to the next stage, so that its effect is identical to a '0' input (the core is not set). However, if this anti-phase winding is connected in series with an in-phase winding from another stage driven by the same timing waveforms, the '1' output from the in-phase winding may be cancelled by the $-$'1' output from the anti-phase winding. In this case the resultant input to the next stage is effectively '0' and no digit will be set. If the two stages are called N (in-phase) and M (anti-phase), where N and M can take the value 0 or 1, a series connection as described above results in the analogue sum $N - M$. The four possible solutions of this sum are -1, 0, 0, and $+1$. The last corresponds to the case where N gives a '1' output while M gives a '0' output and is the only solution that leads to a digit input to the following stage. Thus the analogue sum $N - M$ corresponds logically to the Boolean expression N *and not* $M(N \,.\, \overline{M})$.

This case is shown schematically in Fig. 3(*c*). In the upper part of the diagram the appropriate series connection of the two output windings is illustrated. Below this is shown a useful schematic representation, in which a circle is used to represent one basic unit, and the set and reset times are shown on the left and right respectively. All inputs to one unit are combined by a series connection of output windings from the appropriate stages. An input terminated by an arrow-head indicates an in-phase connection of the corresponding output winding. This will contribute either 0 or $+1$ to the analogue sum at the input. An input terminated by a circle represents an anti-phase connection and contributes either 0 or -1 to the analogue sum. Double arrow-heads and circles are also used and contribute in the former case either 0 or $+2$ and in the latter case either 0 or -2. An input designated 1 is present at every set period (input period); '1' generation is described later. Fig. 3(*a*) shows how a two-input 'and'-gate is constructed. The analogue sum at the input is $N + M - 1$. This takes the value $+1$ only when N and M both contain a digit, and in this case a digit is set in the stage. In Boolean terms the sum is $N \,.\, M$.

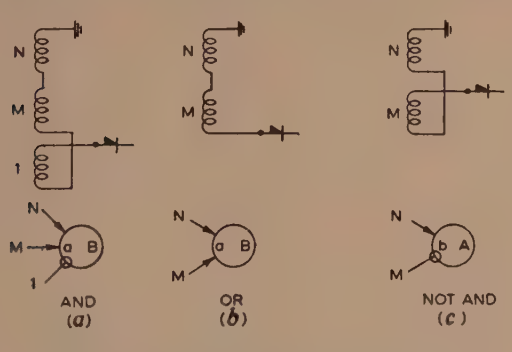


Fig. 3.—The simple logical functions.

In addition to the three simple logical functions shown in Fig. 3, more complex functions may be generated. The analogue sum produced may have, in general, any integral value. A digit will only be set in a stage when this input sum is $+1$ or more. Analogue logic of this type can always be represented by a corresponding Boolean expression, but, since this procedure is usually more complex, it is not recommended. When performing analogue addition, as described above, it is essential for all inputs to be identical in time. This is ensured in the new technique by the strobing action of the setting waveforms, as described in Section 2.1.

While there are in fact four timing waveforms, they may be reduced to two for logical purposes. Thus if the input time, corresponding to the setting time, is *a*, the output time will be *B*. Such an output is connected to a unit having input time *b*. *B* and *b* may be taken logically to be the same time, as may *A* and *a*.

(2.3) Circuit Performance and Tolerance

By observing the following precautions, the performance of the circuit can be made uncritical. All circuit tolerances are to some extent interdependent, but the main effects of each can be described separately.

(2.3.1) Windings.

The serially connected drive windings have many of the properties of a lumped-constant delay line, each section consisting of a series inductance and a parallel capacitance, the latter being principally the stray capacitance between windings on the same core. Thus the drive current applied to a particular core is subject to both amplitude and time distortion determined by the position of the core in the line. Such distortion may be minimized by reducing either the series inductance or the stray capacitance. The former can only be achieved by using large drive-current amplitudes, in which case the required drive-winding turns for a given m.m.f. are correspondingly small. This solution is not always convenient in practice. The capacitance is mainly the total stray capacitance between the drive winding and two other windings—the control winding and the other drive winding. To minimize the stray capacitance, the control winding and the two drive windings are wound independently, each occupying only a small part of the core circumference.

The 2 mm (outside diameter) store cores used at 500 kc/s have a 50-turn control winding which is most conveniently wound as a double strand linking the core 25 times. The output windings use four strands linking the core 10 times.

(2.3.2) Voltage Supplies.

The 10-volt negative supply is used only to provide a biasing current for the transistor, to permit wide variation; with the other parameters (voltage, drive-pulse amplitude and timing) set to their mean value, this voltage may be doubled or halved with little effect.

The positive supply would appear to be more critical, since it defines both the output levels and the time necessary to reset the core completely. In fact, however, the absolute value of the output level is not important, since all outputs will vary proportionately and the input circuit will accommodate a considerable change in amplitude before operation is impaired. To reset the core completely in the available time, the positive supply should be maintained above a minimum voltage level. Its maximum value is determined by the criterion that the output pulse should persist to the end of the set period of the following stage, i.e. the flux should continue to change during this time. In practice, neither of these limits need be adhered to rigidly, since the transfer characteristic[2,7] is such that a long '0' or a short '1' input (the amplitude is the same) can be accommodated. For example, if the positive supply voltage is below the minimum value mentioned above, a core containing a '1' will not be completely reset in the time available. Thus after the passage of a '1' or a series of 1's through a stage, the extra flux stored in the core will lead to an increase in the length of the first '0' output. This effect is offset in practice by choosing the circuit conditions such that any '0' output shorter than half the length of a full '1' output is attenuated by subsequent stages. For the same circuit conditions, '1' outputs more than half the length of a full '1' output will lengthen in passing through subsequent stages. In one application, in which many stages were connected, a variation in the positive supply of from 3·6 to 7·6 volts was permissible.

(2.3.3) Drive-Pulse Amplitude.

The reset-pulse amplitude must exceed a certain minimum value (F ampere-turns) to reverse the core state. The m.m.f. by which this minimum value is exceeded is available for output. Where no output is required the excess is exactly opposed by current drawn via the control winding and diode D3 (Fig. 1), so that only the current-carrying capacity of the diode limits the reset-pulse amplitude. Consequently, the m.m.f. available for output can be made very large, and the circuit has the important property that the information in one stage can be transferred to any number of similar stages, up to the limit set by the diode D3 and by the practical difficulty of providing a large reset m.m.f. In practice, 'branching' to a large number is rarely required, so that the reset m.m.f. is made only large enough to give a 'branching factor' of about 10.

In special cases where a higher figure is required the reset m.m.f. can be increased by employing a stage output winding in series with the normal reset winding.

The set-pulse amplitude must have a similar minimum value to reverse the core state completely. A maximum value is set by the requirement that the transistor, when conducting, must be able to supply sufficient opposing current to prevent flux change in the core. In practice, the set-pulse amplitude is chosen to be half-way between these limits to allow the maximum tolerance, and to permit load current of either sign to be accommodated (Section 2.5).

(2.3.4) Timing.

The permissible tolerances outlined above are all affected in some degree by the timing waveforms. Since these are easily derived from a basic sinusoidal waveform, it is relatively simple to define them accurately. The circuit can accommodate a timing variation of roughly $\pm 10\%$ if necessary.

(2.3.5) Output Variation.

Independent variations of a particular output voltage can occur for the following reasons:

(*a*) Variation of catching-diode characteristic (D3 in Fig. 1).

(*b*) Variation in coupling between control and output windings in different cores.

Since all cores are identically wound, and the toroidal structure leads to almost perfect coupling, the effect of (*b*) is negligible.

Variation due to (*a*) is also small. Allowing a wide range of from 0·2 to 0·6 volt for the diode when conducting leads to an independent voltage variation across the control winding of from 4·7 to 5·1 volts (i.e. only $\pm 4\%$). Thus many output windings may be balanced with accuracy.

The standard unit will accept inputs that depart considerably from their nominal value. A nominal-'1' input, having a value 30% less than that of a true '1', is permissible. Similarly a nominally-zero input may be up to 30% of a true '1' input without causing error. Thus, for example, a series connection of nine windings from different stages, five in-phase and four in anti-phase, is permissible. When all nine stages contain a '1', the analogue sum is nominally +1, but if all outputs are in error by up to $\pm 3\%$ (a practical figure), the sum may be up to 27% low (though the probability of an error of this order is extremely small).

(2.4) Transistor Requirements

Three transistor parameters are of primary importance in the present circuit. One of them, the peak current-carrying capacity, is only important in so far as it affects the earthed-base cut-off frequency, f_{co}, of the transistor, which falls off considerably above a certain current level. Thus f_{co} must be maintained up to the required peak collector current. The latter is primarily a function of the core parameters and the loading requirements of a single stage (Section 3). The third important parameter is the permissible peak collector voltage, which should be as high as possible (see Section 3).

It has been found in practice that a transistor having $f_{co} > 5$ Mc/s is suitable for use at 500 kc/s provided that collector current does not exceed base current by more than a factor of five. An approximate quantitative explanation of this requirement can be made in the following way. The transistor is required to provide current through the control winding to oppose a set pulse completely, i.e. to prevent any flux change in the core. Thus with the collector potential only a little below that of the emitter the transistor is suddenly called upon to provide a pulse of collector current. The amplitude of this current, I_c, is several times that of the base biasing current, I_b, and increases as the set-pulse amplitude is increased. It might be thought that the provision of a biasing current to the base would enable the transistor to provide this transient collector current more readily, since the transistor is, in effect, 'primed'. This is in fact the case when the transistor is maintained in a conducting state for a sufficiently long time. With the input negative, or disconnected, large collector currents can be drawn from the transistor (almost to the value determined by the base current and corresponding d.c. base-to-collector current gain), without the collector potential departing significantly from that of the emitter. However, as normally used, the input is connected to the output of a previous stage, and the transistor is rendered non-conducting at every input period, both '1' and '0' inputs being of sufficient amplitude to ensure this; the '0' input is short, so that the transistor begins to resume its conducting state before the set pulse is applied. Thus to oppose the set pulse, collector current must be provided rapidly by a transistor which is non-conducting, or, at best, partially conducting. With a transistor initially non-conducting, it may be shown[8] that collector current rises approximately to a value I_c in time t_r according to the relation

$$I_c = I_b\beta[1 - \exp(-2\pi f'_{co}t_r)] \quad . \quad . \quad . \quad (1)$$

where f'_{co} is the common-emitter cut-off frequency.

When t_r is very large the exponential term disappears, so that β is the d.c. base-to-collector current gain. In the case of interest, $t_r \ll (2\pi f'_{co})^{-1}$ so that

$$I_c \simeq I_b \beta 2\pi f'_{co} t_r$$

or

$$t_r = \frac{I_c}{I_b} \frac{1}{2\pi} \frac{1}{\beta f'_{co}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Note that the condition for this to be true may be rewritten, using eqn. (2), as

$$\frac{I_c}{I_b} \ll \beta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

i.e. the collector current must rise only to a fraction of the value it would reach under d.c. conditions.

Eqn. (2) shows that the rise time of the collector current is inversely proportional to the product $\beta f'_{co}$. This term is almost exactly equal to f_{co}, the common-base cut-off frequency. Thus finally

$$t_r = \frac{I_c}{I_b} \frac{1}{2\pi} \frac{1}{f_{co}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Taking $I_c/I_b = 5$, eqn. (4) shows that $t_r < 0{\cdot}16$ microsec for $f_{co} > 5$ Mc/s. The practical case differs from the above in two important respects. First, the collector potential is very close to that of the emitter, and secondly the transistor may be partially conducting at the commencement. However, the agreement between theory and practice is satisfactory. Both lead to the conclusion that a transistor having a common-base cut-off frequency greater than 5 Mc/s will provide base-to-collector current gain of at least five, while performing satisfactorily in the circuit under discussion at 500 kc/s.

(2.5) Output Loading

When an input terminal is driven positive, current must be provided as previously explained. In general, this current passes through a number of serially connected output windings. Since such output windings may be connected either in-phase or anti-phase, the current drawn through them may either assist or oppose the driving waveform acting at the same time. This waveform may be either the set or the reset pulse, since an analogue output may be positive in either of these periods (but not in both). Thus the loading on a particular stage resulting from the use of one output winding may be of any one of the four types listed below; in all cases it is assumed that the output winding is connected in series with several others such that the input of the following stage is driven positive.

(*a*) Output winding in-phase. Output current opposes the reset waveform.

(*b*) Output winding in-phase. Output current assists the set waveform.

(*c*) Output winding anti-phase. Output current opposes the set waveform.

(*d*) Output winding anti-phase. Output current assists the reset waveform.

If the resultant input to a stage is negative or zero, current through any output winding connected to that stage is blocked by the input diode. However, it should be noted that the presence of a negative input in one half-cycle (set or reset) must be accompanied by a positive input in the other half-cycle (reset or set). Because of this property, the load type may be determined by a knowledge of two conditions, the phase of the output winding and the analogue sum at the input *at the input time*. This is summarized in Table 1.

Since the analogue sum at the input will in general differ in each digit period, the number and type of output loads on each stage will be similarly variable. However, by applying Table 1 to the logical truth table (as instanced for a full adder in Section 7), the maximum number of each load type may be determined. The limits to this number are calculated in Section 3 and do not prove to be very restrictive in practice.

Table 1

		Analogue sum at input to following stage		
		$\geqslant 1$	0	$\leqslant -1$
Connection of output winding	In-phase	Type *a*	Zero	Type *b*
	Anti-phase	Type *d*	Zero	Type *c*

The other restriction to the logical use of the element is due to the limited voltage excursion allowed at the input terminal. This limit is set in the negative direction by the reverse breakdown voltage of the input diode, and in the positive direction by the maximum permissible reverse voltage between emitter and base of the transistor. The former is not an important restriction in practice, since it is typically -100 volts, while the transistor limit is about $+12$ volts for a good transistor. A positive input of $+12$ volts cannot be accompanied by a negative input of more than about -24 volts (assuming the peak voltage output from the core is almost double the mean voltage). The positive voltage restriction at the input limits the analogue sum at the input to about 10 (allowing a little over one volt for the output voltage level). Thus any 'or' gate up to a 10-input 'or' gate may be constructed. More inputs may be handled by providing two or more diodes at the input. Similarly, a negative limitation of about -7 must be placed on the analogue sum at the input, since a negative input in the input period is accompanied by a positive input in the other half-cycle.

(3) DESIGN

Generalized results are derived on the basis of the circuit of Fig. 1.

To ensure adequate discrimination between the two binary outputs, it is necessary for the voltage integral permitted across the control winding during the set period, with the transistor conducting, to be small compared with that for complete reversal. Thus if v_0 is the voltage developed across the diode D2 and the saturated transistor, we must have

$$v_0 t \ll \frac{\Phi_s C}{100} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The number of control-winding turns, C, is chosen to be as large as possible, to ensure the above inequality, and to achieve a high stage power gain, as explained later. The upper limit is set by the maximum permissible negative voltage that can be tolerated at the transistor collector, and by practical winding considerations.

Since a core containing a '1' must be reset to '0' in a time T microseconds, we have

$$v = \frac{\Phi_s C}{100T} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

If the required input voltage to a unit is v_{in}, the output turns are determined by

$$n = \frac{C v_{in}}{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

For a base current of I_b amperes, the transistor can supply collector current of

$$\hat{I}_c = \beta_{max} I_b \text{ amperes} \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where β_{max} is the maximum base-to-collector current gain that can be employed.

Thus the total set m.m.f. must be limited to

$$\hat{F}_1 = \hat{I}_c C + F_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where F_0 is the m.m.f. required to drive the core at the restricted rate.

The minimum set m.m.f. must be

$$\check{F}_1 = F \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where F is the m.m.f. required to reverse the core state completely in t microseconds. Thus if F_{in} and F_{out} are the set and reset m.m.f.'s applied to the core respectively, the maximum number of output loads that may be driven are

$$y_a = \frac{F_{out} - F}{nI_b} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

$$y_b = \frac{\hat{F}_1 - F_{in}}{nI_b} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$y_c = \frac{F_{in} - \check{F}_1}{nI_b} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

y_d is limited only by the diode (Section 2.3.3).

In a practical case, the limits would be set somewhat lower to allow for initial transient loading and for reasonable drive-amplitude tolerances. For the design of Fig. 1, employing a 2 mm (outside diameter) store core at 500 kc/s, $y_b = y_c = 10$ for an input m.m.f. of 2 AT, while $y_a = 20$ for an output m.m.f. of 3 AT.

y_a and y_d are not affected by the transistor characteristics and may be adjusted, in theory, to have any value; y_b and y_c can also be varied, but not independently. It is readily shown from the above that

$$y_b + y_c = \frac{v\beta_{max}}{v_{in}} - \frac{(F - F_0)\beta_{max}}{n\hat{I}_c} \quad . \quad . \quad . \quad (14)$$

This expression may be taken to be the stage power gain. The second term may be ignored if $\hat{I}_c$ is made large enough, and the power gain of the circuit is thus seen to be a product of the voltage gain and the transistor current gain. However, since F must increase with the rate of reversal of the core, it may not be possible at high digit rates to increase $\hat{I}_c$ sufficiently to prevent some reduction of power gain.

(4) 1's GENERATORS AND CONDITIONAL-1's GENERATORS

A 1's generator is quite simply derived from the standard unit (Fig. 1) by omitting the transistor and associated diodes, D1 and D2, as shown in Fig. 4. Thus two components only are necessary, the core and diode D3 producing a standard output at each digit period.

A conditional-1's generator consists of a standard unit controlled by a d.c. input which is either greater than +1 volt or negative. Control may be from a hand switch or a staticizor.

(5) WHOLE-DIGIT DELAY

The addition of a capacitor to the basic unit as shown in Fig. 5 provides temporary storage of the input pulse, and enables a delay of a full digit period to be obtained economically. The input-pulse amplitude is somewhat larger than that used for the basic unit and is derived from the centre-tap of the control winding. A '1' input charges the storage capacitor and drives the base of the transistor positive. The transistor is held non-conducting for the duration of the input pulse and subsequently while the capacitor is discharged. During the discharge, set pulse *b* drives the core to the '1' condition. A '0' input is considerably shorter, only partially charging the storage capacitor. The latter is discharged sufficiently to allow the transistor to resume its conducting state before the onset of set pulse *b*, so that no digit is set in the core.

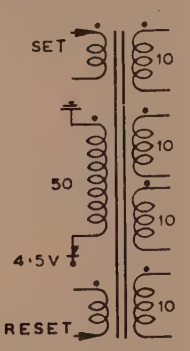


Fig. 4.—1's generator.

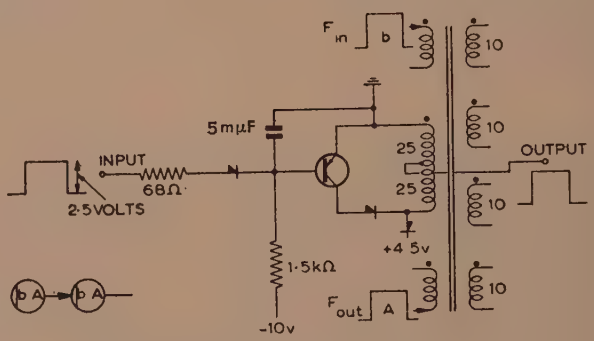


Fig. 5.—Whole-digit delay unit at 500 kc/s.

While proper design can ensure the above mode of operation, the circuit of Fig. 5 is less useful than the standard circuit (Fig. 1). In the standard circuit the '0' input is ignored and complex analogue addition can be performed at the input, since the set pulse acts as a strobe on the inputs as previously explained. The whole-digit delay unit has neither of these properties and, in addition, can only accept inputs of a given amplitude. (Double-amplitude inputs charge the input capacitor further so that discharge is not complete before the next digit period commences.) Use of the unit might be justified in equipment containing a large number of shift registers. The equivalent logical symbol (Fig. 5) has a bar at the input to indicate the additional half-digit delay.

(6) TIMING WAVEFORM GENERATOR

The timing waveform generator (Fig. 6) employed when working at a digit rate of 500 kc/s makes use of a sine-wave generator transformer-coupled to three phase-shifting networks, each of which is followed by a squaring stage. Two of these are variable and are used for the set waveforms. The third is fixed and provides a square wave for the two reset waveforms. The outputs are coupled to conventional driver circuits.

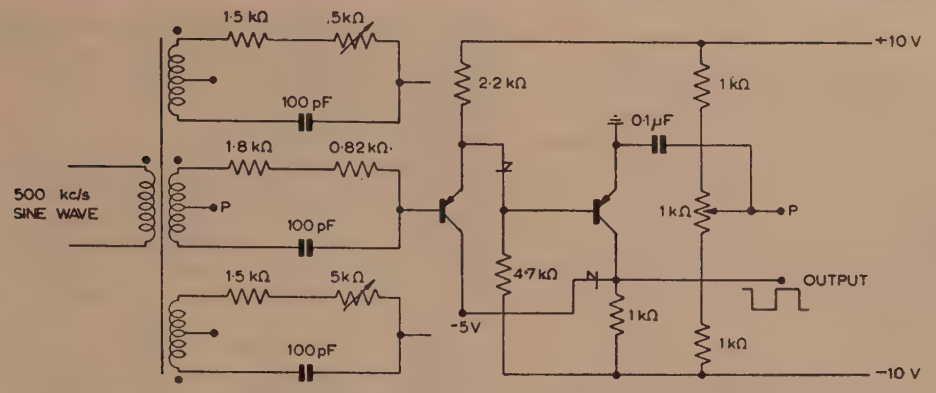


Fig. 6.—Timing waveform generator (500 kc/s).

(7) AN ADDER

Fig. 7 shows schematically a 3-input adder employing three elements. Only the elements C′, X and Y are necessary to produce the sum and carry. When the adder is to be used for the serial addition of two inputs, a fourth stage C″ is required to provide an additional half-digit delay for the carry. The output of C″ is fed back to the input of the adder in this case. Table 2 contains the logical truth table. In addition, output loading is compiled for each stage as an illustration. The fifth stage in Fig. 7 is not part of the adder and is included only to show the analogue addition performed at the output of the previous stages. Reference to Table 2 will show that the analogue sum at the input to this stage is identical to the logical sum at the output, except for the half-digit-time delay. Since the input to the first three elements contains the common term A + B + C, only one output winding from each of stages A, B and C is required.

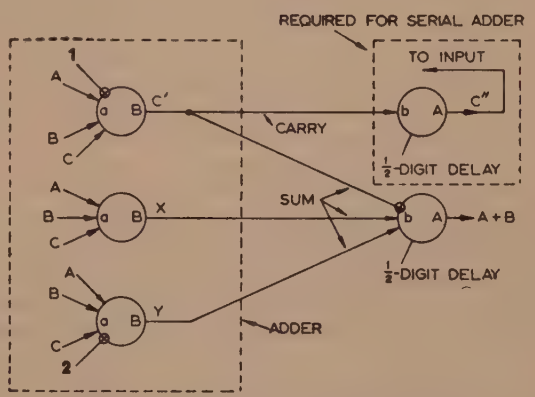


Fig. 7.—Schematic of a 3-input adder.

(8) A TEST SYSTEM

Fig. 8 shows a simple block diagram of a chain-code generator which produces, repetitively, a binary pattern 127 digits in length. The pattern is generated by detecting the contents of the last two stages of a seven-stage shift register. Where this is similar, a '1' is inserted at the front of the register in the next digit period; where different, a '0' is inserted. The resulting sequence is such that the content of the register includes every possible combination of seven binary digits in each cycle, with one exception: the condition in which all seven stages contain a '1' does not occur.

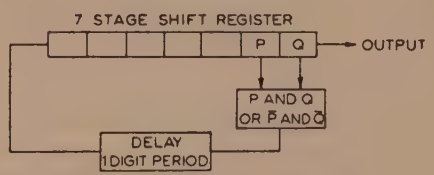


Fig. 8.—Chain code generator.

Table 2

TRUTH TABLE AND LOADING FOR A 3-INPUT ADDER

A B C	C′ =A+B+C−1		Load	X =A+B+C		Load	Y =A+B+C−2		Load	Total load	A+B =X+Y−C′		Load		C″ =C′	Load	Total load
	L.S.	A.S.	A, B, & C	L.S.	A.S.	A, B, & C	L.S.	A.S.	A, B, & C	A, B, & C	L.S.	A.S.	X + Y	C′	L.S. A.S.	C′	C′
0 0 0	0	−1	*b*	0	0	0	0	−2	*b*	2*b*	0	0	0	0	0	0	0
0 1 0	0	0	0	1	1	*a*	0	−1	*b*	*a* + *b*	1	1	*a*	*d*	0	0	*d*
1 0 0	0	0	0	1	1	*a*	0	−1	*b*	*a* + *b*	1	1	*a*	*d*	0	0	*d*
1 1 0	1	1	*a*	1	2	*a*	0	0	0	2*a*	0	0	0	0	1	*a*	*a*
0 0 1	0	0	0	1	1	*a*	0	−1	*b*	*a* + *b*	1	1	*a*	*d*	0	0	*d*
0 1 1	1	1	*a*	1	2	*a*	0	0	0	2*a*	0	0	0	0	1	*a*	*a*
1 0 1	1	1	*a*	1	2	*a*	0	0	0	2*a*	0	0	0	0	1	*a*	*a*
1 1 1	1	2	*a*	1	3	*a*	1	1	*a*	3*a*	1	1	*a*	*d*	1	*a*	*d* + *u*
									Peak load →	3*a* + 2*b*	↑ Sum		*a*		↑ Carry		*d* + *u*

L.S. = Logical sum at output.
A.S. = Analogue sum at input.

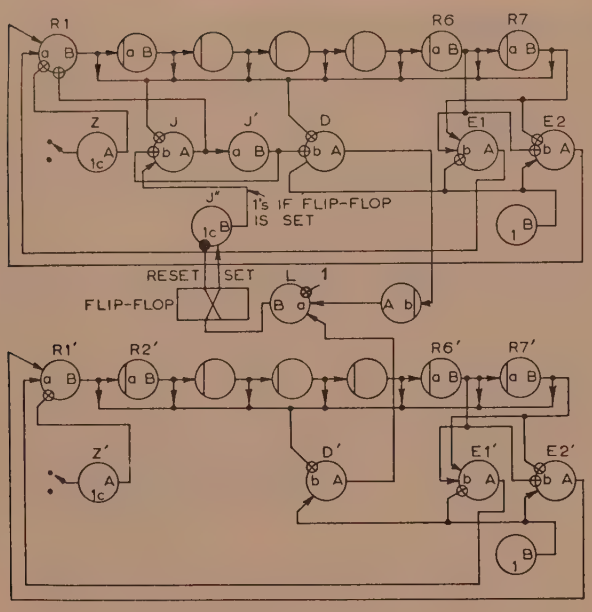


Fig. 9.—Test system.

The system shown in Fig. 9 consists basically of two chain-code generators. The conditional-1's generators Z and Z′ are controlled by hand switches and serve to reset the patterns to the 'all 0' condition as required. D and D′ detect the presence of 'all-0's' in the shift register. The same condition is detected by J only if the flip-flop controlling the conditional-1's generator J″ is set. The flip-flop may be set as required, but is reset if the two patterns are in phase. To show how the two patterns may be brought into phase, consider the flip-flop to be initially set. When the pattern reaches the all-0 condition, J gives an output which inhibits the '1' input to the register from E2 and insets a '1' into J′. The register is still in the all-0 condition, so that the pattern has been effectively delayed by one digit period. However, J does not give a second output since input to this stage is inhibited by the output from J′. Thus the stage E2 is able to insert a '1' at the beginning of the register. Thereafter, the pattern proceeds on its normal course until the all-0 condition repeats and the pattern is again delayed by one digit period. Eventually, the all-0 condition occurs for both generators together. The 'and' gate L detects this point and resets the flip-flop, preventing further phase change between the patterns. The patterns continue in step until one of them is reset to zero.

It is worth noting that the input to J consists of a series connection of nine output windings. In spite of this, and of the extensive use of delay stages, the margin of tolerance on this circuit in practice has been surprisingly good.

(9) CONCLUSIONS

A digital technique has been described that fulfils most of the requirements of a logical building block, as set out in the Introduction. While the number of components used in each unit is fairly high, the high speed and logical flexibility of the resulting system gives adequate compensation. The use of a square-loop ferrite at 500 kc/s necessarily involves a fairly high power level, though this compares favourably with techniques using cores alone. Further improvement may result from the use of tape-wound cores requiring a smaller applied m.m.f. for a given switching time. At higher speeds, power consumption is increased proportionately.

With readily available components, the circuit operates reliably at 500 kc/s. To achieve higher speeds, three factors must be considered. First, the core losses associated with continuous reversal of the core state must not cause the Curie temperature to be approached. At present available data are insufficient to determine the necessary upper limit to the reversal rate for a particular core assembly. Secondly, if the stage power gain [eqn. (14)] is to be maintained at higher digit rates, the requirements of the transistor become more severe. Eqn. (4) shows that if the rise time of the transistor collector current is to be constrained to a small fraction of the digit period, operation at higher speeds requires the use of a transistor having a higher cut-off frequency. Further, the reduced voltage and current rating of high-frequency transistors make the achievement of a high power gain more difficult. Finally, it follows that a limited increase in speed is possible with the same components if a lower stage power gain is acceptable.

(10) ACKNOWLEDGMENTS

The work described was carried out under a development contract awarded by the Signals Research and Development Establishment to Ferranti Ltd. The author's thanks are due to the Director of the S.R.D.E. for permission to publish the paper, and to Mr. P. H. Cutler, also of the S.R.D.E., who suggested the test system described in Section 8.

(11) REFERENCES

(1) GUTERMAN, S., KODIS, R. D., and RUHMAN, S.: 'Circuits to Perform Logical and Control Functions with Magnetic Cores', Convention Record of the Institute of Radio Engineers, 1954, Part 4, p. 124.

(2) SCARROTT, G. G., HARWOOD, W. J., and JOHNSON, K. C.: 'The Design and Use of Logical Devices using Saturable Magnetic Cores', *Proceedings I.E.E.*, Paper No. 2065 M, March, 1956 (**103** B, Supplement No. 2, p. 302).

(3) KARNAUGH, M.: 'Pulse-Switching Circuits using Magnetic Cores', *Proceedings of the Institute of Radio Engineers*, 1955, **43**, p. 570.

(4) LOEV, D., MIEHLE, W., PAIVINEN, J., and WYLEN, J.: 'Magnetic Core Circuits for Digital Data-Processing Systems', *ibid.*, 1956, **44**, p. 154.

(5) GUTERMAN, S. S., and CAREY, W. M.: 'A Transistor-Magnetic Core Circuit: A New Device applied to Digital Computing Techniques', Convention Record of the Institute of Radio Engineers, 1955, Part 4, p. 83.

(6) PERRY, G. H., HOFFMAN, G. R., and SHALLOW, E. W.: 'A New and Simple Type of Digital Circuit Technique using Junction Transistors and Magnetic Cores', *Proceedings I.E.E.*, Paper No. 2112 M, November, 1956 (**103** B, Supplement No. 3, p. 412).

(7) ELDRIDGE, D., and DOREY, P. F.: 'A Core-Transistor Logical Element', *ibid.*, Paper No. 2347 R, July, 1957 (**104** B, Supplement No. 7, p. 512).

(8) MOLL, J. L.: 'Large-Signal Transient Response of Junction Transistors', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1773.

DISCUSSION ON THE ABOVE TWO PAPERS BEFORE THE MEASUREMENT AND CONTROL SECTION, 4TH NOVEMBER, 1958

Dr. G. B. B. Chaplin: The paper by Mr. Barnes and Dr. Stephen is the first in this country to present operating experience on a transistor computer, and so the authors' comments are especially interesting. Point-contact transistors have always had a reputation for unreliability, but the authors have shown that, apart from initial troubles, these transistors are at least as reliable as valves. The figure of over 90% serviceable for two months is really remarkable for a computer left unattended overnight and at weekends.

I note that a substitute circuit has been designed which uses two junction transistors, and I feel it significant that, although the point-contact transistor it replaces was first produced in about 1952, and six years of unprecedented effort have since gone into developing junction transistors, the authors had difficulty in finding junction transistors good enough to simulate the point-contact transistor. Even then they had to resort to selection to get sufficiently high collector voltages.

The authors quote some failures of the junction transistor, and I should like to know if these tended to occur fairly soon after the transistors were inserted. If so, the failure rate should improve with time.

An interesting feature of Mr. Eldridge's circuit is that outputs taken during the 'reset' period [i.e. (*a*) and (*d*)] can be very large, since the core is operating like a magnetic amplifier. However, it seems that the resulting high gain cannot be used to advantage in practice, because outputs taken during the 'set' period are limited by the transistor gain. Does Mr. Eldridge consider this a fair statement, and, if so, would it be possible—at the expense, perhaps, of some circuit duplication—to arrange that outputs are not taken during the 'set' period?

Some time ago we had occasion at Harwell to design a basic computing element using high-frequency transistors, the whole emphasis being on speed of operation. We decided to use transformers to allow analogue addition of voltages, as Mr. Eldridge has done, but were not able to use square-loop material because of the speed required. In most other respects the systems were very similar, the transistor acting as a switch which was closed just before the clock pulses were applied. A fundamental limitation which emerged was that, as the number of windings in series is increased, the operating speed of the circuit must be decreased. In normal diode gating, for example, the correct output is usually obtained by the time the input pulses have risen to about two-thirds of their final value, but if nine analogue voltages are being added, the output pulse must rise to well over 90% of its final value, and this may take a very long time if the leading edge approaches its final voltage exponentially. In fact, we gave up stacking windings for this reason, but still used transformers because they allowed matching of circuits and inversion of waveforms. Did Mr. Eldridge have to take this limitation into account when designing his circuit?

Dr. R. L. Grimsdale: A transistor digital computer* was completed at Manchester University in 1955, but a prototype was running in November, 1953. This computer, like the one described by Mr. Barnes and Dr. Stephen, used a magnetic-drum store and point-contact transistors, and the circuits operated at the (then) high speed of 125 kc/s. The computer was operated for six months, and during that time it was observed that, on the average, trouble was experienced with two of the 200 transistors each week. The trouble appeared to be a gradual deterioration of the characteristics and, in particular, an increase in I_{co}.

All the transistors received from the manufacturers were tested, and a noticeable improvement was observed in those delivered later in 1953 as compared with those received earlier in the year. All were of the unsealed type: do the authors consider that a reliable point-contact transistor could have been produced by using the improved techniques of manufacture and hermetic sealing now employed for high-speed junction transistors? It seems to be generally accepted that the point-contact transistor is obsolete, but this is to be regretted because a 2-state circuit can be made with one point-contact transistor, making it possible to construct a computer with very few transistors.

Mr. J. A. Bailey: In 1955 my firm decided to proceed with the development of a point-contact-transistor computer, in consultation with Dr. Grimsdale, and we too found that such transistors require selection before use. This same difficulty seems to appear with junction transistors.

The computer contained approximately 300 point-contact transistors, together with a number of valves. On testing transistors as received from the manufacturers, about 10% were rejected for use in our particular circuit and a further 20% failed when the machine was first switched on. Once testing was complete, the failure rate fell to about 2¼% per month and later to about 1% per month. It was noted that the failure rate was somewhat higher in winter than in summer. Most failures were located on commencing work in the morning, the machine being switched off at night. We consider that this was possibly due to thermal cycling and contamination of the material of the transistor.

For a machine commissioned at the beginning of this year we have taken a 40-hour week as the normal operating time. The maintenance and development curves show a peak in the middle, due to drum trouble, when the machine had to be shut down, and to further development. We were not worried about the time taken, since we had two other machines in operation at the time. The failure rate of transistors fell to approximately one per month in August and September. In August, 1957, it was decided to convert this machine to use junction transistors, the approximate replacement rate being 3½ : 1; with 1 100 transistors in the machine we can calculate this as 1 344 000 transistor-hours in nine months, during which we had three failures, all internal and mechanical. We have since had no failures at all. In comparison, with point-contact transistors we had 52 failures in the same period. Have Mr. Barnes and Dr. Stephen noticed a particularly higher rate of failure due to leaving the machine switched off for some time?

Mr. J. P. Bunt: One of the difficulties when designing a logical circuit is to choose its capabilities. The adder shown by Mr. Eldridge is neat, but adders are not typical of logical devices. The test is, what does the logical designer do? The author shows three basic functions: 'and', 'or' and 'not and'. The 'and' circuit seems expensive in −1 sources and, in practice, logical functions can be produced adequately by using the 'or' and 'not and' exclusively. Compared with diode logic, one can do less per stage, but this is compensated for by having two stages for every digit delay.

Mr. Barnes and Dr. Stephen say that their machine is marginally tested for 30 min a day; that this is a valuable aid in bringing all circuits up to a uniform standard of reliability; and that waveforms are examined every two months. There seems to be a contradiction here. Marginal testing is useful to improve design values and detect initial mistakes, but it is surprising that a daily test is done with transistor circuits, since the frequency of marginal testing should be related to the rate of change of the

* KILBURN, T., GRIMSDALE, R. L., and WEBB, D. C.: 'A Transistor Digital Computer with a Magnetic-Drum Store', *Proceedings I.E.E.*, Paper No. 2043 M, March, 1956 (103 B, Supplement No. 3, p. 390).

parameters. Surely a weekly or monthly test would be adequate if the appropriate margins are employed?

I note that the minimum acceptable condition for using the machine is that an error-free run of at least 30 min under marginal conditions is performed. If the appropriate margins are being applied to the machine, an isolated error in 30 min would suggest that under normal conditions the machine would be reliable. If this is not so, the margins which are being applied are not suitably chosen.

Mr. E. H. Cooke-Yarborough: I wish to add a footnote to the paper by Mr. Barnes and Dr. Stephen. Both authors went on leave at the same time this summer, but we did not shut the machine down; we left it computing and detailed someone to keep it fed with cards. Without any other attention it computed successfully for a fortnight, and was still at work when the authors returned.

The figures which the authors have given for reliability relate to last year, and it would be interesting to have more up-to-date figures.

In the written paper by Mr. Eldridge the bias on the catching diode of Figs. 1 and 5 is shown as 4½ volts, whereas on the lantern slides it is 3·6 volts. Was this change made to overcome some difficulty experienced with the original circuits?

The author mentions a power consumption figure of ½ watt per logical unit. How much of this is dissipated in the core? Does the core temperature rise appreciably?

Section 2.4, dealing with transistor requirements, does not mention the reverse breakdown voltage between the emitter and the base. It would appear that, when rather complex logical functions are being performed, there is a possibility of the base being driven several volts positive with respect to the emitter. This may exceed the rating of some types of transistor.

Mr. W. S. Elliott: In listening to the paper by Mr. Barnes and Dr. Stephen I was reminded that it is nearly 11 years since I heard in this lecture theatre a paper on a digital computer. So far as I can remember, this is the first critical paper dealing with operating experience on a computer since then. We have heard many computers described, especially at the Convention on the subject which was held here two years ago. It is a great pity that we have not previously had any papers on operating experience, and I hope that in the future we shall have more papers such as this one.

Mr. A. C. Lynch: I agree that Mr. Eldridge's algebra is not Boolean, but surely it is still digital? Why use the word 'analogue'?

Mr. A. Peacock: I have used Mr. Eldridge's logical device in a large-scale demonstration model and found it an extremely easy and flexible element with which to design. The model handles binary-coded decimal and alphabetical data in the serial mode with one parity check bit per character. Both the decimal and alphabetical bits are taken through the arithmetic unit, which can add, subtract, automatically recomplement and detect overflows. Arithmetic is checked by a separate calculation of the correct parity change. The entire unit contains about 50 logical elements, of which eight supply timing pulses and six are 1's generators.

This gives the practical answer to the suggestion that the system would require rather a large number of 1's generators. Five-input 'and' gates would require four 'inhibit' windings from a 1's generator; but the gates in a serial system are more commonly only 2-input, or 3-input, 'and' gates. It is also possible to wire each gate with the 1's generator at the earthed end and hence use a single winding to drive several gates.

Dr. J. R. Tillman (*communicated*): I hope this paper will be the forerunner of others describing field, or near-field, experiences with transistors. Its value would have been increased if more quantitative data had been given, (*a*) of the initial values of the key parameters of the transistors, because type numbers sometimes imply wide spreads, (*b*) of the criteria for failure, (*c*) of the performance of those which 'failed', at the time of failure, t_f (their performance at say $0{\cdot}9t_f$ in addition is perhaps too much to expect) and (*d*)—although this is not so conveniently obtained—of the performance of a sample at one or two times during the year of use. Failing these quantitative data, we run the risk of a wide divergence of views on reliability, arising in part because different designers allow different margins of deterioration beyond the limits of acceptable initial performance.

THE AUTHORS' REPLIES TO THE ABOVE DISCUSSION

Mr. R. C. M. Barnes and **Dr. J. H. Stephen** (*in reply*): Since the paper was prepared we have accumulated a further year's operating experience with the computer. The reliability record for the first 10 months of 1958 is shown in Fig. A, and it will be noticed that the performance has improved over that shown in Fig. 1 of the paper, except for one month—July, 1958—when the mechanical adjustments of the magnetic drum were disturbed. Between September, 1957, and October, 1958, one junction transistor and eight point-contact transistors failed (seven of type TP1 and one of type EW61). The overall failures per 1 000 hours are now approximately 0·3% for junction transistors, 0·8% for TP1 unsealed point-contact transistors and 0·4% for EW61 sealed point-contact transistors. We were disappointed when the EW61 transistor ceased to be available, since it seemed to be well on the way towards the robust hermetically-sealed component which could have redeemed the reputation of the point-contact transistor. Our figures for failures do not give a true picture of the merits of the EW61 compared with earlier point-contact transistors, particularly as regards its ability to withstand handling, its freedom from failures on the shelf and when the computer is switched off, and a pleasing absence of the fickle faults which had given the point-contact transistor a bad name. As Mr. Bailey has also found, the TP1 transistor is very liable to develop tedious faults when the computer is switched off, and this is one reason why we have tried to keep Cadet running continuously.

We have not experienced any failures early in the life of junction transistors. Our few failures have been confined to 3X/300N and 3X/301N transistors, which have been in use for several years. When modifying the reading amplifiers of the magnetic-drum store we introduced 55 transistors (types GET12, GET13 and OC72) which have now been operating for a year without failure. It is probable, however, that larger changes of transistor parameters can be tolerated in the amplifiers than in logical elements.

We agree with Mr. Bunt that the frequency of marginal tests should be related to the rate of change of parameters. Unfortunately, the older point-contact transistors in Cadet suffer from sudden and erratic changes of parameters, so that frequent testing is desirable. The magnetic-drum store also tends to develop sudden faults, owing to changes in the mechanical adjustments, and this is a further reason for daily testing. The routine examination of waveforms was discontinued early in 1958. When a 30 min test produces only one error it is very probable, as Mr. Bunt suggests, that the computer will be error-free for the next day under normal conditions, but we have often found that the isolated error is a useful warning of an incipient fault. Although the relative amplitudes of the alternating

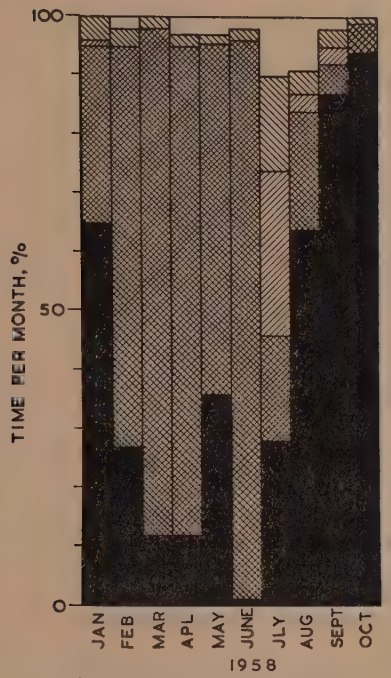


Fig. A.—Serviceability of Cadet since December, 1957.
The key given in Fig. 1 of the paper still applies.

marginal test voltages might be improved, we are satisfied that our present test conditions are a reasonably sensitive indicator of the state of the computer.

We fully agree that we should have collected the data which Dr. Tillman mentions. The experimental model of Cadet was originally intended as a temporary measure while we assembled a tidy, compact version of the computer, and this, together with our experience that TP1 transistors were easily damaged when handled, deterred us from making regular measurements of transistor parameters, apart from an initial inspection when each batch arrived from the manufacturer. We should like to emphasize Dr. Tillman's remarks in the hope that others will be encouraged to carry out this useful work.

Our criteria for transistor failure are, first, that the transistor shall have misbehaved in the computer, and secondly, that it shall then either be outside the initial acceptance limits ($I_{co} < 2$ mA for $|V_{b-c}| = 15$ volts and $I_e = 0$; and $|V_{b-c}| < 2$ volts for $I_c = 5{\cdot}5$ mA and $I_e = 3$ mA) or shall show the floating-emitter effect. There may well be transistors which are operating reliably although they are now outside the acceptance limits. It would be interesting to conclude the experiment by testing all the transistors in the computer when it reaches the end of its useful life.

Mr. D. Eldridge (*in reply*): Dr. Chaplin makes a fair statement of the circuit gain limitation, although it is a little misleading to refer to outputs taken during the 'set' period. Such outputs do limit the circuit as it is described in the paper, but are entirely unwanted. There are a number of ways in which this unwanted output could be avoided, but the most attractive would be to return the output connections, not to earth, but to a voltage waveform which swings sufficiently negative during the 'set' period to avoid any net positive input during this time. However, this added complication, together with the practical difficulty of providing a large 'reset' pulse, does not seem justified in practice, where loading limitations are not usually troublesome. It is also true, as Dr. Chaplin suggests, that unwanted outputs could be avoided logically, but the logical restrictions would be serious. Thus, all gates which could produce a negative analogue sum would be inadmissible.

No difficulty has been experienced in connecting many windings in series, except for voltage loss across the winding resistance which limits a series connection to 15 windings, or less if tappings are used. This difference in experience may be due to

(*a*) Different circuit-impedance levels.
(*b*) The fact that in the present circuit only analogue sums of +1 or 0 are significant.
(*c*) The use of a catching diode and a large 'reset' current, which enables the core output to achieve its final value rapidly.

Mr. Cooke-Yarborough mentions a difference between the voltage level shown on a slide at the presentation and that given in the paper. This arises from a more recent attempt to reduce the peak collector voltage applied to the transistor in accordance with manufacturers' ratings (by reducing the control winding turns), although no difficulty had been encountered in practice.

Reverse emitter-base voltage limitations are mentioned in Section 2.5, but it is worth noting that short transient reverse voltages considerably in excess of the normal d.c. rating may be applied to the base of a transistor.

The power dissipation of the core is, of course, a function of the digit pattern passing through a stage, but where '1's are continuously passed the core dissipates approximately 150 mW.

621.317.39

The Institution of Electrical Engineers
Paper No. 2801 M
Mar. 1959

CIRCUITS EMPLOYED IN THE N.P.L. CAESIUM STANDARD

By L. ESSEN, O.B.E., D.Sc., Ph.D,, Associate Member, E. G. HOPE, B.Sc., and J. V. L. PARRY, M.Sc.

(*The paper was first received* 18*th June, and in revised form* 16*th October,* 1958.)

SUMMARY

A description is given of the electronic circuits used for the excitation of the caesium resonance and the measurement of the resonant frequency, attention being drawn to the precautions which have to be taken in order to obtain an accuracy of ± 1 part in 10^{10}. The exciting source is derived by the multiplication of the frequency 5·006 9 Mc/s of a quartz oscillator. The oscillator is then set to give the resonant frequency exactly. The measurements are made at the high frequency of approximately 9 Gc/s so that an accuracy of 1 c/s is obtained in one second using a counter in the final stage.

An alternative and rather simpler system is also described, and some comments are made on the problem of frequency synthesis.

(1) INTRODUCTION

The description of the N.P.L. caesium standard given in a previous paper[1] dealt in detail with the atomic-beam chamber, but only in broad outline with the electronic circuits used for exciting the atomic resonance at a frequency of 9 192 631 830 c/s and for relating this to the 100 kc/s standard quartz oscillators. The design of these circuits is important if the quartz clocks are to be calibrated in terms of the caesium standard with a precision of 1 part in 10^{10}, and the present paper contains a description of those used successfully at the National Physical Laboratory.

Caesium standards can be used in two ways. An oscillator can be set at the value corresponding to the spectral line, as determined by a maximum in the number of atoms received at a detector, and can then be used to calibrate the quartz standards in terms of the caesium frequency. Alternatively, the oscillations exciting the atomic resonance can be derived by synthesis from a quartz oscillator, and the atomic-beam signal then used in a servo loop to maintain the quartz-oscillator frequency in a fixed relationship to that of caesium. Only the former method has so far been used at the N.P.L. because there were some advantages in using the caesium standard first as a passive resonator and thus avoiding the complications and possible errors of a servo mechanism. This approach is well suited to a standardizing institution already equipped with quartz clocks but would be less convenient to some users. Whichever method is used, considerable latitude is possible in the detailed arrangement of equipment, and the choice made at the N.P.L. was to some extent governed by what was available at the time. After three years' experience, it is felt that the system is essentially sound, although some simplification would be possible. A description is therefore also given of a simpler equipment which has been constructed and on which preliminary tests have been carried out.

It will be convenient to deal with the equipment on a functional basis. Throughout the description, details are given only of those features which are particularly important, the circuits simply calling for sound engineering practice being given only in schematic form.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
The paper is an official communication from the National Physical Laboratory.

(2) EXCITATION OF THE ATOMIC RESONANCE

(2.1) General Requirements

The requirements for the exciting oscillator are freedom from sidebands, a bandwidth of not more than a fraction of 1 c/s, a stability of 1 c/s during the period occupied by the setting and measurement, and a range of frequency variation adequate to scan the spectral line and the neighbouring lines in the Zeeman pattern. The power required was not known at the beginning of the investigation and it was decided to provide about 30 mW, which can be obtained from the ordinary local-oscillator type of klystron. It was found later that the power required to induce satisfactory transitions is of the order of 100 μW.

The first precise measurement of the frequency gave the value 9 192 631 831 c/s at a field of 0·05 oersted, corresponding to a value of 9 192 631 830 c/s at zero field. This value has been used provisionally at the N.P.L. to define an atomic unit of time interval, although it is appreciated that it is based only on the value of the astronomical unit made available at the time of the measurements (June, 1955). A value of 9 192 631 770 $\pm$ 20 c/s has now been obtained[2] in terms of the second of Ephemeris Time averaged over the period June, 1955, to June, 1958. This value will be used in the future.

(2.2) The Pound Controlled Oscillator

Work at the N.P.L.[3] has shown that under suitable conditions an oscillator controlled by a high-Q cavity resonator in the manner described by Pound could be made to fulfil the above conditions. The model already used in several applications had been designed with the caesium work in mind and was centred on a frequency of 9 192 Mc/s with a fine frequency adjustment of ± 5 Mc/s. It was used successfully for the early part of the investigation and for routine measurements for about one year. The experience demonstrated its suitability for microwave spectroscopy even for the very high resolution of 1 part in 10^{10}, but the circuit adjustments required to achieve the high stability and narrow bandwidth were extremely critical and could be tedious. For work at a single frequency, or over a frequency band of a few parts in 10^5, it is more convenient to use a high harmonic of a quartz oscillator. As soon as the frequency and power requirements were known, therefore, multipliers for a quartz oscillator were constructed.

(2.3) Quartz Harmonic Source

There were several reasons for not attempting to derive the exciting source directly from one of the 100 kc/s quartz-ring standards available. A complicated process of synthesis would be required to give the exact spectral-line frequency, and experience had shown that it is difficult to synthesize from a frequency as low as 100 kc/s without introducing phase modulation. It is, moreover, against general practice to adjust the frequency of the standards even by the small amount necessary to scan the line. It was therefore decided to begin with a crystal of known good performance at a frequency close to 5 Mc/s. The crystal was adjusted to 5·006 9 Mc/s and had a fine control for varying the frequency by the necessary amount. The multiplier is shown in

Fig. 1, in which the frequencies of the various stages are given to the first five figures.

An output from the crystal oscillator at 10·013 Mc/s (a doubler is incorporated) is multiplied to 170·23 Mc/s by the selection, in a valve harmonic generator, of the 17th harmonic. This signal is then amplified in two preliminary stages including four tuned circuits to give a power of about 0·5 watt, which is amplified to an available power of 5 watts by three further stages. There are altogether eleven circuits tuned to 170·23 Mc/s, and the level of the nearest 10·013 Mc/s sideband is less than −60 dB relative to the wanted signal. Sufficient power is taken from this amplifier to drive a type QQV03–20A valve operating as a frequency trebler which gives a maximum power output of about 2 watts at 510·70 Mc/s and drives a further trebler, consisting of a type CV273 valve in a non-contact coaxial-line earthed-grid circuit with untuned cathode input, followed by a similar coaxial-line-amplifier stage with tuned input and output circuits. The power from this unit at 1 532·1 Mc/s is between 0·5 and 1 watt. The frequency is then doubled to 3 064·2 Mc/s by an EC57 disc-seal triode with an untuned cathode input circuit and an anode circuit consisting of a tuned cavity coupled by a probe to a short length of WG10 waveguide (Fig. 2). A second probe mounted across the guide is connected directly to a germanium diode, type GEX66, in a WG16 waveguide at right angles to and immediately above the larger guide. The No. 16 guide is coupled through an attenuator to the atomic resonator. The power available in the smaller guide at the third harmonic of the EC57 output frequency is approximately 1 mW, which is ample for the excitation of the atomic transitions.

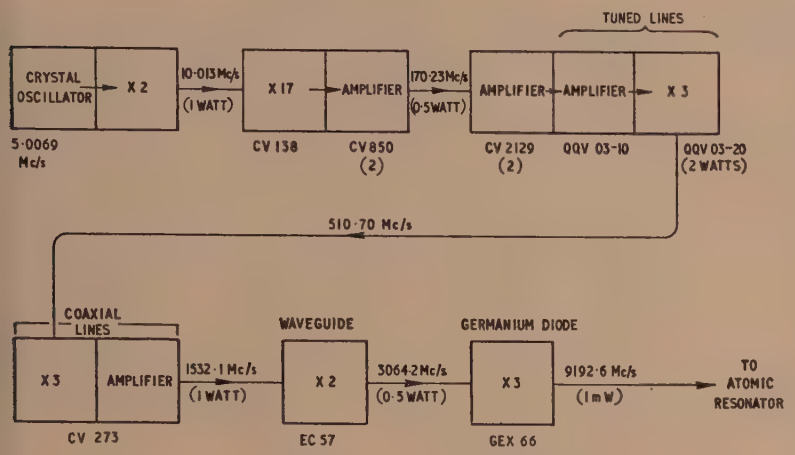


Fig. 1.—Energizing source for caesium resonator.

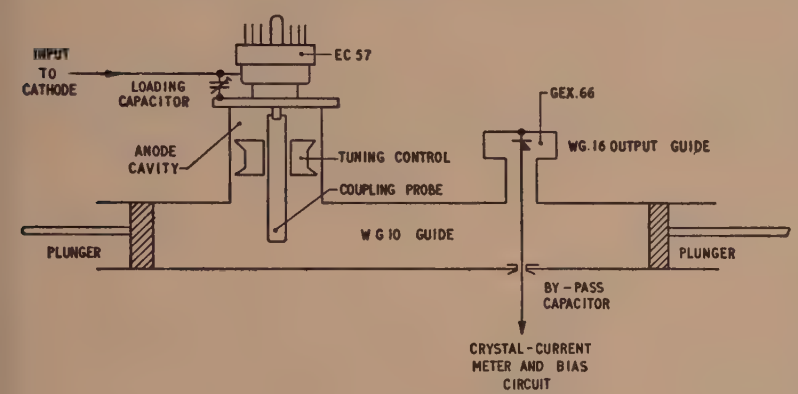


Fig. 2.—EC57 waveguide multiplier.

(3) FREQUENCY MEASUREMENT

The method of calibrating the 100 kc/s quartz standard shown diagrammatically in Fig. 3 is simple and gives the required accuracy of 1 part in 10^{10} in a measuring time of one second. The exciting oscillation described in Section 2.3 is set to the caesium resonance at a frequency of 9 192 631 831 c/s and is mixed in a germanium diode with a stronger signal (0·25 watt) at a nominal frequency of 200 Mc/s derived from a 5 Mc/s oscillator. From the mixer is taken the difference frequency between the exciting

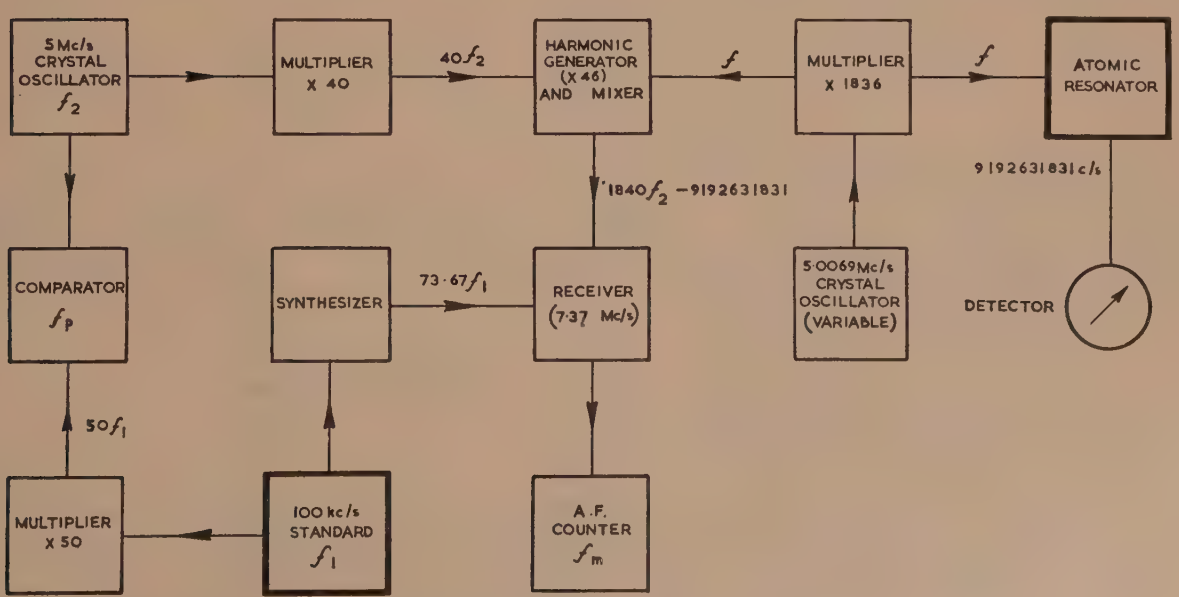


Fig. 3.—Frequency measuring system for caesium resonator.

oscillation, f(c/s), and the 46th harmonic of the 200 Mc/s signal. If the exact frequency of the 5 Mc/s oscillator is f_2(c/s) the difference frequency is

$$1\,840f_2 - 9\,192\,631\,831 \simeq 7{\cdot}368 \times 10^6 \text{ cycles per second} \quad (1)$$

This frequency is in turn mixed in a receiver with a signal synthesized from the 100 kc/s (nominal) standard. If the frequency of

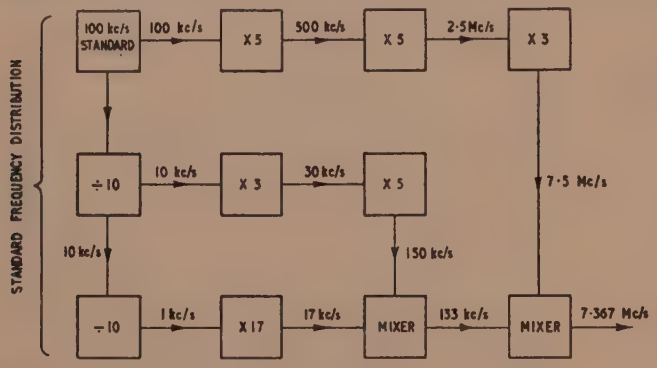


Fig. 4.—Frequency synthesizer for 7·367 Mc/s.

this standard is f_1(c/s), the synthesized frequency is $73{\cdot}67f_1$ cycles per second. The output frequency of the receiver is then

$$f_m = 1\,840f_2 - 9\,192\,631\,831 - 73{\cdot}67f_1 \quad . \quad . \quad . \quad (2)$$

This residual frequency, of the order of 1 kc/s, is measured on a counter either directly or by adjusting a separate audio-frequency oscillator to equality with it and then measuring the

Schematics of the synthesizer and receiver are given in Figs. 4 and 5.

(4) ALTERNATIVE SYSTEMS OF EXCITATION AND MEASUREMENT

A simpler scheme (Fig. 6) which has recently been tried in experimental form eliminates one of the multiplier chains and the synthesizer. The exciting source is derived from a 5 Mc/s instead of a 5·006 9 Mc/s quartz crystal, the small difference in frequency now being produced by adding at the 170 Mc/s stage the output from an auxiliary quartz oscillator of approximately 234 kc/s which is variable over a small range. The resultant frequency of 170·234 Mc/s is the same as that formerly produced by direct multiplication from 5·006 9 Mc/s and the subsequent stages are unchanged. In order to avoid the difficulty of separating the signals of 170 ± 0·234 Mc/s, the mixing is done in two stages, the 234 kc/s being first added to the 10 Mc/s output from the 5 Mc/s crystal (Fig. 7), and the resultant then being added to 160 Mc/s (Fig. 8). The 160 Mc/s could clearly be obtained by four doubling stages from 10 Mc/s, but in the present equipment the existing circuits are retained and merely retuned from 170 Mc/s to 160 Mc/s.

A separate circuit is used to measure the precise frequency of the auxiliary quartz oscillator when set to give the peak of the caesium resonance (Fig. 9). A signal at 233·3 kc/s is produced from the 100 kc/s standard and a beat in the audio range is obtained by mixing this with the crystal frequency. The audio beat is multiplied by 32 before being counted so as to give the required accuracy of 1×10^{-10} in a time of about 2 seconds.

This alternative system, as described, still requires the fre-

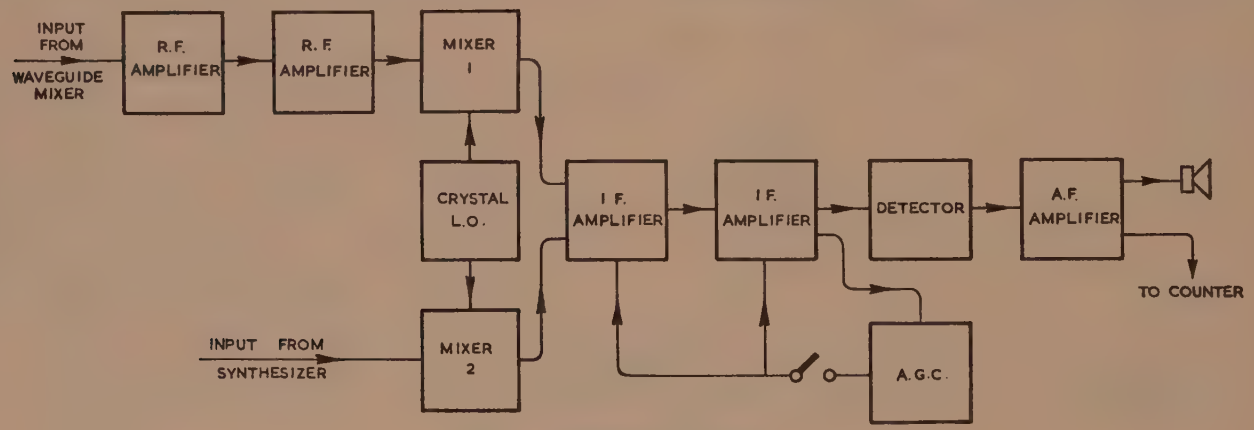


Fig. 5.—Receiver for 7·367 Mc/s.

a.f. oscillator on the counter. With the latter method there is no need to monitor the level and waveform of the signal from the receiver.

Throughout the measurement, the frequency difference between the 5 Mc/s oscillator and the 50th harmonic of the 100 kc/s standard is measured, giving

$$f_p = 50\,f_1 - f_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Eliminating f_2 from eqns. (2) and (3), we obtain the value of f_1 in terms of the caesium resonant frequency as

$$f_1 = \frac{9\,192\,631\,831 + f_m + 1\,840f_p}{92\,000 - 73{\cdot}67} \text{ cycles per second}$$

The fractional deviation from the nominal value is

$$\frac{f_1 - 10^5}{10^5} = \frac{f_m + 1\,840f_p - 1\,169}{0{\cdot}919\,3} \times 10^{-10}$$

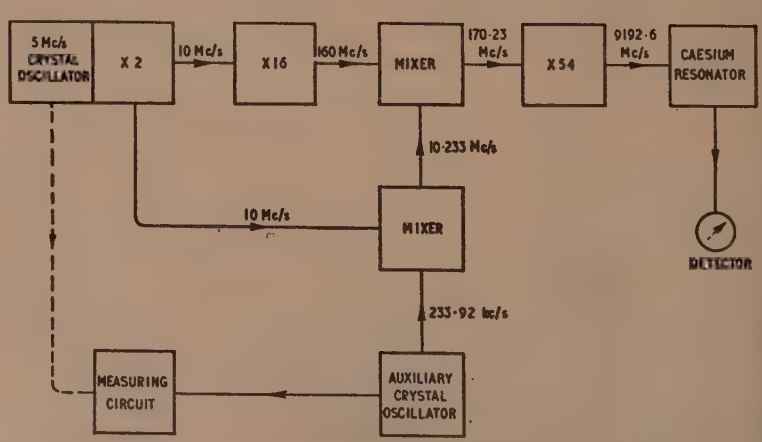


Fig. 6.—Energizing source using auxiliary crystal oscillator.

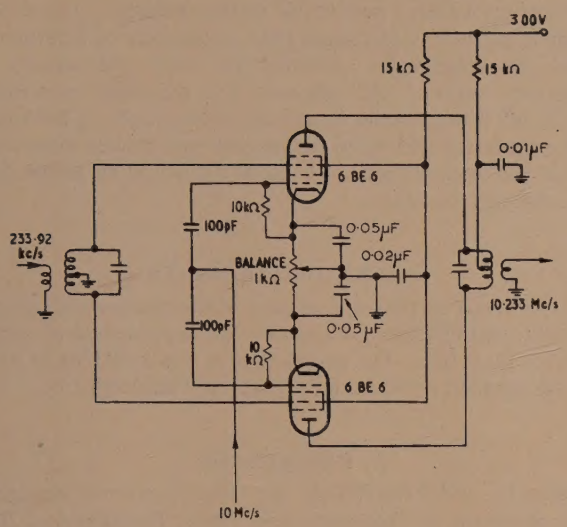


Fig. 7.—Mixing stage for 10·233 92 Mc/s.

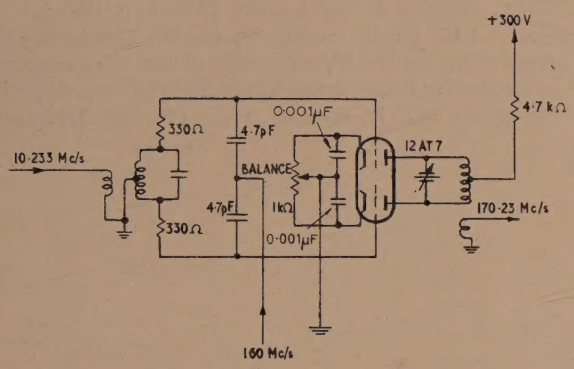


Fig. 8.—Mixing stage for 170·234 Mc/s.

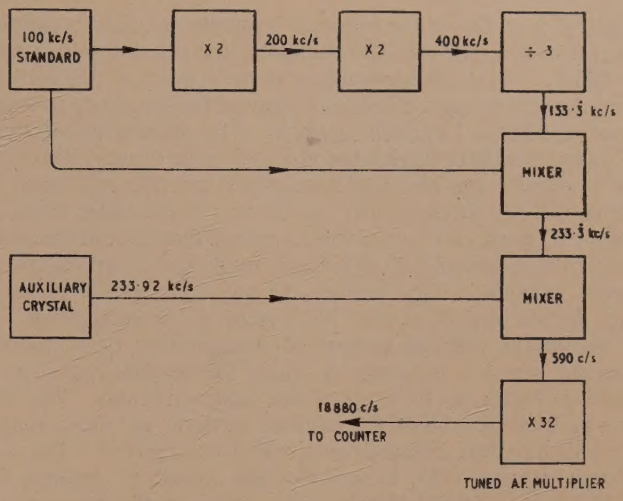


Fig. 9.—Auxiliary crystal measuring circuit.

quency of the 5 Mc/s oscillator to be measured in terms of the 100 kc/s standard. It would be convenient to eliminate this step by constructing a multiplier chain operating from the 100 kc/s standard and providing an output at 5 Mc/s to replace the existing crystal oscillator. Experiments with several different designs of multiplier have shown, however, that even with d.c. heater and highly stabilized h.t. supplies throughout, it is difficult to limit phase modulation to the low value required to achieve a stability of 1 part in 10^{10} in the output. Further experiments are continuing in order to determine whether such a scheme is practicable.

(5) SYNTHESIS OF THE CAESIUM FREQUENCY

If the caesium resonance is used to control the frequency of a quartz oscillator through a servo loop it is necessary to synthesize the resonant frequency from the quartz frequency. There are two basic methods of doing this, although there is a further choice of multiplying and dividing factors and methods of combination in order to achieve the correct relationship. One basic

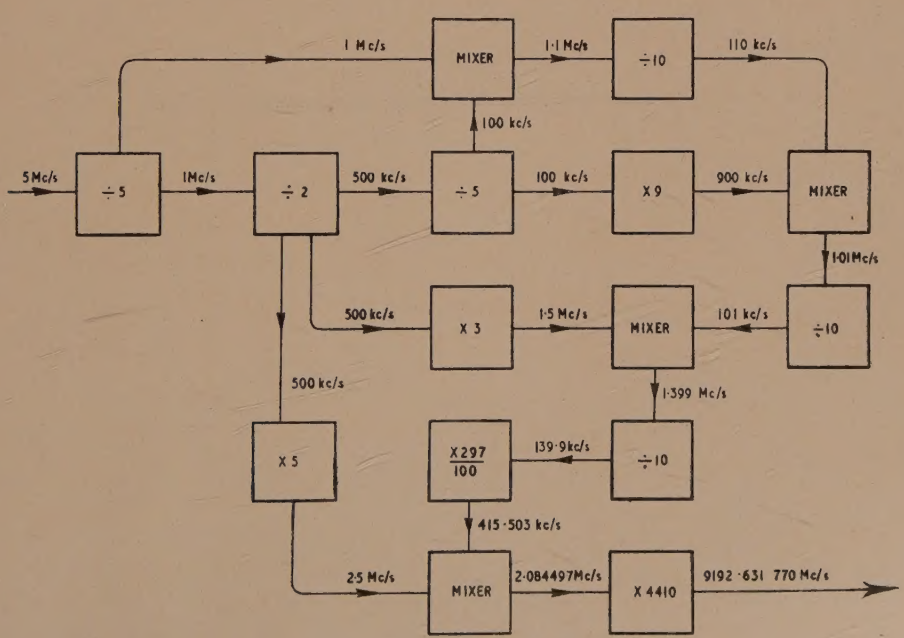


Fig. 10.—Synthesizer for 9 192 631 770 c/s.

method is to derive the caesium frequency by direct multiplication from a quartz oscillator adjusted to a suitable value such as 5·0069 . . . Mc/s, to apply the servo mechanism to this oscillator and then construct a circuit to transpose the frequency to a value such as 5 Mc/s, 1 Mc/s or 100 kc/s. The other is to begin with a quartz oscillator having the nominal value of, say, 5 Mc/s and to synthesize the caesium frequency from this and apply the servo directly to the quartz oscillator. If the latter method is adopted, great care is needed to ensure that a pure frequency, free from sidebands of significant amplitude, is applied to the beam chamber. This implies the reduction of all unwanted signals to a level of at least 40 dB below the wanted signal.

Before the detailed process of transposition or synthesis is worked out it is necessary to know the adopted value of the caesium frequency in terms of the second of time. The value can be ascertained with the required accuracy only by averaging the astronomical measurements over a long period. The value found in June, 1955, in terms of the second of uniform time (UT2) was 9 192 631 830 c/s. The value at the beginning of 1958 was, however, more nearly 9 192 631 930 c/s, indicating that the unit of time has changed. There is the further complication that the unit of time now adopted by the International Committee of Weights and Measures is the second of Ephemeris Time and the value obtained in this unit is 9 192 631 770 $\pm$ 20 c/s. A synthesizer giving this value is shown in Fig. 10. It has not been constructed, but the design takes into account all the technical requirements. Whatever value is finally chosen, it should be possible to design a fairly simple process of synthesis to obtain a relationship within 1 part in 10^9 of that desired. If a different number is used, it is of course only a question of arithmetic to convert the frequencies obtained by using the standard to the correct values. At present, for example, a value of 9 192 631 840 is used in the American Atomichron, and allowance must be made for this when comparing frequencies measured in terms of the Atomichron with those measured in terms of the N.P.L. caesium standard.

(6) ACKNOWLEDGMENTS

The work was carried out as part of the research programme of the National Physical Laboratory and is published by permission of the Director. The quartz crystal was an RD6170 BT-cut standard supplied by Marconi's Wireless Telegraph Co.

(7) REFERENCES

(1) ESSEN, L., and PARRY, J. V. L.: 'The Caesium Resonator as a Standard of Frequency and Time', *Philosophical Transactions (of the Royal Society of London)*, *A*, 1957, **250**, p. 45.

(2) MARKOWITZ, W., HALL, R. G., ESSEN, L., and PARRY, J. V. L.: 'Frequency of Caesium in terms of Ephemeris Time', *Physical Review Letters*, 1958, **1**, p. 105.

(3) ESSEN, L.: 'A Highly Stable Microwave Oscillator and its Application to the Measurement of the Spatial Variations of Refractive Index in the Atmosphere', *Proceedings I.E.E.*, Paper No. 1422 R, January, 1953 (**100**, Part III, p. 19).

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING), MARCH 1959

CONTENTS

	PAGE
Rating of Speech Links and Performance of Telephone Networks.....J. SWAFFIELD, Ph.D., B.Sc.(Eng.), and D. L. RICHARDS, B.Sc.(Eng.)	65
Assessment of Speech Communication Links.....D. L. RICHARDS, B.Sc.(Eng.), and J. SWAFFIELD, Ph.D., B.Sc.(Eng.)	77
Discussion on the above two papers	90
The Use of Dispersive Artificial Dielectrics in a Beam-Scanning Prism.....J. S. SEELEY, B.Sc.(Eng.), Ph.D., and J. BROWN, M.A., Ph.D.	93
The Quarter-Wave Matching of Dispersive Materials.....J. S. SEELEY, B.Sc.(Eng.), Ph.D.	103
Theory of Reflections from the Rodded-Type Artificial Dielectric.....A. CARNE, B.A., M.Sc., and J. BROWN, M.A., Ph.D.	107
Discussion on the above three papers	114
An Investigation of the Excitation of Radiation by Surface Waves.....K. P. SHARMA, M.Sc.	116
The Launching of Radial Cylindrical Surface Waves by a Circumferential Slot...J. BROWN, M.A., Ph.D., and K. P. SHARMA, M.Sc., Ph.D.	123
Parallel-Plate Transmission Systems for Microwave Frequencies.....A. F. HARVEY, D.Phil., B.Sc.(Eng.)	129
Optical Techniques at Microwave Frequencies.....A. F. HARVEY, D.Phil., B.Sc.(Eng.)	141
A New Method of Generating a Rotating Radiation Polar Diagram.....H. W. HAWKES, B.Sc.	158
Transistor Active Filters using Twin-T Rejection Networks.....A. E. BACHMANN, Dipl.El.Ing.	170
The Transformation of Admittance through a Matching Section and Lossless Waveguide Junction.....J. R. G. TWISLETON, B.Sc.	175
The Power Radiated by a Surface Wave circulating around a Cylindrical Surface.....Prof. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.)	180
The Recognition of Moving Vehicles by Electronic Means.....T. S. PICK, B.Sc.(Eng.), and A. READMAN	186
Pattern Recognition by means of Automatic Analogue Apparatus.....W. K. TAYLOR, B.Sc., M.Sc., Ph.D.	198
A System for the Automatic Recognition of Patterns. R. L. GRIMSDALE, M.Sc., Ph.D., F. H. SUMNER, B.Sc., Ph.D., C. J. TUNIS, B.Eng., M.Sc., Ph.D., and T. KILBURN, M.A., Ph.D., D.Sc.	210
Operating Experience with a Transistor Digital Computer.....R. C. M. BARNES, B.Sc.(Eng.), and J. H. STEPHEN, Ph.D.	222
A New High-Speed Digital Technique for Computer Use.....D. ELDRIDGE, B.A.	229
Discussion on the above two papers	237
Circuits employed in the N.P.L. Caesium Standard.....L. ESSEN, O.B.E., D.Sc., Ph.D., E. G. HOPE, B.Sc., and J. V. L. PARRY, M.Sc.	240

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) *which are out of print and from which reprints are not available.* The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: 'Reflections from the Ionosphere', *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

★ There are many members and former members of The Institution who are finding life difficult. **When remitting your membership subscription,** please help them by sending a donation, or an annual subscription preferably under deed of covenant, to

THE BENEVOLENT FUND

The object of the Fund is to afford assistance to necessitous members and former members (of any class) of The Institution of Electrical Engineers who have paid their subscriptions for at least five years consecutively or compounded therefor, and to the dependants of such members or former members.

Subscriptions and Donations may be sent by post to

THE INCORPORATED BENEVOLENT FUND OF THE INSTITUTION OF ELECTRICAL ENGINEERS, SAVOY PLACE, LONDON, W.C.2

or may be handed to one of the Local Hon. Treasurers of the Fund.

Though your gift may be small, please do not hesitate to send it

★

Local Hon. Treasurers of the Fund

EAST MIDLAND CENTRE . . . . . *R. C. Woods*
IRISH BRANCH . . . . . *A. Harkin, M.E.*
MERSEY AND NORTH WALES CENTRE. *D. A. Picken*
NORTH-EASTERN CENTRE . . . *J. F. Skipsey, B.Sc.*
NORTH MIDLAND CENTRE *E. C. Walton, Ph.D., B.Eng.*
SHEFFIELD SUB-CENTRE. . . . . *F. Seddon*
NORTH-WESTERN CENTRE . *E. G. Taylor, B.Sc.(Eng.)*
NORTH LANCASHIRE SUB-CENTRE *G. K. Alston, B.Sc.(Eng.)*
NORTHERN IRELAND CENTRE . *G. H. Moir, J.P.*
SCOTTISH CENTRE . . . . . *R. H. Dean, B.Sc.Tech.*
NORTH SCOTLAND SUB-CENTRE . . . . *P. Philip*
SOUTH MIDLAND CENTRE . *Capt. J. H. Patterson, R.A.*
RUGBY SUB-CENTRE . . . . . *P. G. Ross, B.Sc.*
SOUTHERN CENTRE . . . . . *G. D. Arden*
WESTERN CENTRE (BRISTOL) . . . *A. H. McQueen*
WESTERN CENTRE (CARDIFF) . . . *E. W. S. Watt*
WEST WALES (SWANSEA) SUB-CENTRE . *O. J. Mayo*
SOUTH-WESTERN SUB-CENTRE . . *W. E. Johnson*

Members are asked to bring to the notice of the Court of Governors any deserving cases of which they may have knowledge

THE BENEVOLENT FUND

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.